

In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institution shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

THE RELATIONSHIP BETWEEN THE PROPERTIES OF CERTAIN
SAMPLE STATISTICS AND THE STRUCTURE OF ACTIVITY
IN SYSTEMATIC ACTIVITY SAMPLING

A THESIS

Presented to

The Faculty of the Graduate Division

by

William Whaley Hines

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial Engineering

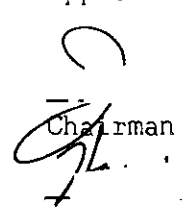
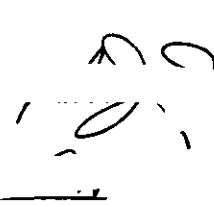
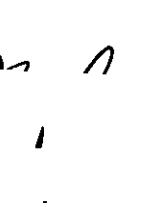

Georgia Institute of Technology

March, 1964

76
12 R

THE RELATIONSHIP BETWEEN THE PROPERTIES OF CERTAIN
SAMPLE STATISTICS AND THE STRUCTURE OF ACTIVITY
IN SYSTEMATIC ACTIVITY SAMPLING

Approved:


Chairman   

Date approved by Chairman: April 6, 1964

ACKNOWLEDGMENTS

The inspiration for this research resulted from an awareness of the gap between the practice of systematic activity sampling and the existing substantiating theory. Dr. Harold O. Davidson introduced the author to this problem, and he provided encouragement in the early stages of the research.

Dr. Joseph J. Moder, the thesis advisor, was most helpful in the direction he gave this research. His comments regarding technical content and organization have been invaluable.

Special credit is also due to Dr. Harrison Wadsworth and Dr. William F. Atchison for their suggestions on the manuscript.

I am appreciative of the aid given by Professor Frank F. Grose-close during the years 1956-1963. I shall always be grateful for the encouragement and assistance from my mother and father during my entire academic career. Finally, I acknowledge the thoughtfulness and encouragement given by my wife, Gayle, and children, William and Jennifer, during the time of my graduate study.

TABLE OF CONTENTS

| | Page |
|---|------|
| ACKNOWLEDGMENTS | ii |
| LIST OF TABLES | vi |
| LIST OF ILLUSTRATIONS | ix |
| SUMMARY | xii |
| Chapter | |
| I. INTRODUCTION | 2 |
| Objectives of Study | |
| Importance of the Study of Systematic Activity Sampling | |
| Definitions | |
| Scope of Study | |
| Study Procedure | |
| Assumption | |
| II. PRESENT STATE OF SYSTEMATIC ACTIVITY SAMPLING | 9 |
| Introduction | |
| The Initial Work | |
| Process Stability | |
| Other Sources of Variation | |
| Groups with Correlated Activity | |
| Other Models | |
| The Development of Activity Sampling | |
| Systematic Activity Sampling | |
| III. A REVIEW OF SYSTEMATIC SAMPLING LITERATURE | 31 |
| Introduction | |
| Early Papers Relative to Systematic Sampling Applications | |
| The Initial Development of Theory | |
| Increased Interest in Systematic Sampling | |
| The Correlogram in Analyzing Systematic Designs | |
| Recent Developments in Systematic Sampling | |
| Concluding Remarks | |
| IV. POPULATION DESCRIPTION AND SYSTEMATIC SAMPLING DESIGNS FOR SIMPLE ACTIVITY STRUCTURE | 54 |
| Introduction | |
| Population Description | |
| Periodicity in the Population | |

| | | |
|-------|--|-----|
| | Infinite Population | |
| | Finite Population | |
| | Systematic Sampling Designs | |
| | One Random Start | |
| | Multiple Random Starts | |
| V. | A SINGLE OCCURRENCE OF THE ACTIVITY | 61 |
| | Introduction | |
| | The Mathematical Model and Sampling Designs | |
| | Comparison of Simple Random and Systematic Designs | |
| | The Distribution of $\hat{p}_{1.sy.d}$ | |
| | Observations Over Several Time Periods | |
| | Concluding Remarks | |
| VI. | MULTIPLE OCCURRENCE OF THE ACTIVITY ON THE PERIOD $[0,T]$ WITH SEVERE RESTRICTIONS ON THE DISTRIBUTIONS OF SPAN LENGTHS . . | 82 |
| | Introduction | |
| | The Mathematical Model | |
| | Estimation and the Error of Estimate | |
| | Concluding Remarks | |
| VII. | MULTIPLE OCCURRENCE OF THE ACTIVITY ON THE PERIOD | 93 |
| | Introduction | |
| | Exponential Distributions of the Span Lengths of A and A' | |
| | Other Distributions of the Span Lengths of A and A' | |
| | Introduction | |
| | Simulation Experiments | |
| | Gamma Distributions of the Span Lengths of A and A' | |
| | Normal Distributions of the Span Lengths of A and A' | |
| | Calculation of Serial Correlations | |
| | Formulation | |
| | Test of Logic and Computer Program for Simulation and Calculation of Serial Correlations | |
| | Results | |
| | Preliminary Analysis | |
| | Secondary Analysis | |
| VIII. | ESTIMATION WITH MULTIPLE OCCURRENCE OF THE ACTIVITY | 139 |
| | Introduction | |
| | One Random Start | |
| | An Upper Bound to the Variance of the Estimate | |
| | With Knowledge of the Process | |
| | With No Knowledge of the Process | |
| | Multiple Random Starts | |
| | Study Over One Period | |
| | Study Over L Periods | |
| | Concluding Remarks | |

| | |
|--|-----|
| IX. STUDY OF NATURAL POPULATIONS | 147 |
| Introduction | |
| Presentation of Data | |
| Analysis | |
| Multiple Random Starts | |
| Conclusions | |
| X. POPULATION DESCRIPTION AND SYSTEMATIC DESIGNS FOR COMPLEX ACTIVITY STRUCTURE | 169 |
| Introduction | |
| Infinite Population | |
| Finite Population | |
| Systematic Sampling Designs | |
| XI. ESTIMATION OF PARAMETERS ASSOCIATED WITH COMPLEX ACTIVITY. . | 175 |
| Introduction | |
| Single Random Start | |
| Multiple Random Starts | |
| Concluding Remarks | |
| XII. CONCLUSIONS AND RECOMMENDATIONS | 179 |
| APPENDIX A | 183 |
| APPENDIX B | 222 |
| LITERATURE CITED | 262 |
| OTHER REFERENCES | 274 |
| VITA | 277 |

LIST OF TABLES

| Table | | Page |
|-------|---|------|
| 1. | Comparison of Simulated and Theoretical Results for Systematic Sampling of a Period on Which the Activity Occurs Once, $T = 480$, $a = 61$, and $p = 0.1271$, 300 Replications for each Value of n | 79 |
| 2. | Probability on Range Quartiles | 89 |
| 3. | Values of Parameters Employed in Simulation with Gamma Distributions for Lengths of Span A and A' | 103 |
| 4. | Values for Parameters Employed in Simulation with Truncated Normal Distributions for Lengths of Span A and A' | |
| 5. | Actual vs. Predicted Values for Gamma Densities | 129 |
| 6. | Predicted vs. Actual values for y_s for Truncated Normal Distributions | 130 |
| 7. | Possible Systematic Samples | 132 |
| 8-A. | Average and Maximum Values of $Q(k)$ for Gamma Distributed Span Lengths | 134 |
| 8-B. | Average and Maximum Values of $Q(k)$ for Span Lengths with Truncated Normal Distributions | 135 |
| 9. | Summary of Calculations from Natural Populations | 150 |
| 10. | Per Cent of Confidence Intervals Containing the Mean. | 166 |
| 11. | Computer Program for Simulated Sampling Where the Activity Occurs Once | 223 |
| 12. | Simulation of a Process Where Span Lengths Have Gamma Densities | 224 |
| 13. | Simulation of a Process Where Span Lengths Have Truncated Normal Densities | 225 |
| 14. | Program for Calculation of the Correlogram | 226 |
| 15. | Simulation of a Process Where the Span Lengths Have Exponential Densities | 227 |

| | | |
|-----|---|-----|
| 16. | Analysis of Simulation Results | 228 |
| 17. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$, $RN = 1190000001$, and $C = 17^{11}$ | 229 |
| 18. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$, $RN = 2559066073$, and $C = 5^{13}$ | 230 |
| 19. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$, $RN = 4999465853$, and $C = 5634765623$ | 230 |
| 20. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$, and Random Number Input | 231 |
| 21. | Values of the Correlation Function from Simulation for $\lambda = 0.4$, $\lambda' = 0.08$, $r = 2$, $r' = 2$ | 231 |
| 22. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 1.0$, and $r = r' = 10$ | 232 |
| 23. | Values of the Correlation Function from Simulation for $\lambda = 3.0$, $\lambda' = 2.0$, $r = 9$, $r' = 14$ | 232 |
| 24. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 4.0$, $r = r' = 16$ | 233 |
| 25. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 2.0$, $r = r' = 20$ | 233 |
| 26. | Values of the Correlation Function from Simulation for $\lambda = 5.0$, $\lambda' = 4.0$, $r = 30$, $r' = 16$ | 234 |
| 27. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 4.0$, $r = 6.4$, $r' = 16$ | 234 |
| 28. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 10.0$, $r = 50$, $r' = 30$ | 235 |
| 29. | Values of the Correlation Function from Simulation for $\lambda = \lambda' = 5.0$, $r = r' = 50$ | 235 |
| 30. | Values of the Correlation Function from Simulation for $\mu = \mu' = 5.0$, $\sigma^2 = \sigma'^2 = 6.0$ | 236 |
| 31. | Values of the Correlation Function from Simulation for $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 6.0$ | 236 |

| | | |
|-----|--|-----|
| 32. | Values of the Correlation Function from Simulation for $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 4.0$ | 237 |
| 33. | Values of the Correlation Function from Simulation for $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 2.0$ | 237 |
| 34. | Values of the Correlation Function from Simulation for $\mu = 4.0$, $\mu' = 20.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 4.0$ | 238 |
| 35. | Values of the Correlation Function from Simulation for $\mu = 30.0$, $\mu' = 5.0$, $\sigma^2 = \sigma'^2 = 4.0$ | 238 |
| 36. | Values of the Correlation Function from Simulation for $\mu = \mu' = 10.0$, $\sigma^2 = 2.0$, $\sigma'^2 = 0.0$ | 239 |
| 37. | Values of the Correlation Function from Simulation for $\mu = 5.0$, $\mu' = 30.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 4.0$ | 239 |
| 38. | Values of the Correlation Function from Simulation for $\mu = 4.0$, $\mu' = 20.0$, $\sigma^2 = \sigma'^2 = 0.25$ | 240 |
| 39. | Analysis of Natural Population | 241 |
| 40. | Extended Analysis of Simulation Data | 242 |
| 41. | Multiple Random Starts - Simulation | 248 |
| 42. | Data from Study No. 11 (Span Durations in Minutes) . . | 254 |
| 43. | Data from Study No. 12 (Span Durations in Minutes) . . | 255 |
| 44. | Data from Study No. 13 (Span Durations in Minutes) . . | 256 |
| 45. | Data from Study No. 14 (Span Durations in Minutes) . . | 257 |
| 46. | Data from Study No. 15 (Span Durations in Minutes) . . | 258 |
| 47. | Data from Study No. 16 (Span Durations in Minutes) . . | 258 |
| 48. | Data from Study No. 21 (Span Durations in Minutes) . . | 259 |
| 49. | Data from Study No. 22 (Span Durations in Minutes) . . | 259 |
| 50. | Data from Study No. 23 (Span Durations in Minutes) . . | 260 |
| 51. | Data from Study No. 24 (Span Durations in Minutes) . . | 261 |

LIST OF ILLUSTRATIONS

| Figure | | Page |
|--------|--|------|
| 1. | Error Associated with Systematic Sampling | 27 |
| 2. | Simple Activity Structure | 55 |
| 3. | The Division of a Period into Intervals | 63 |
| 4. | Sampling Variances for Different Methods of Sampling . | 75 |
| 5. | The Distribution of $\hat{p}_{1 \text{ sy } d}$ for $kd < a < (k+1)d$; $k = 0, 1, \dots, (n-1)$ | 77 |
| 6. | Correlation Functions from Simulated Process Realiza- tions, and Theoretical Values, Where $\lambda = \lambda' = 0$, and $r = r' = 1$ | 95 |
| 7. | Correlogram from Simulation; $\lambda = 0.08$, $\lambda' = 0.4$, $r = r' = 2$ | 108 |
| 8. | Correlogram from Simulation; $\lambda = \lambda' = 1.0$, $r = r' =$ 10 | 109 |
| 9. | Correlogram from Simulation; $\lambda = 3.0$, $\lambda' = 2.0$, $r = 9$, $r' = 14$ | 110 |
| 10. | Correlogram from Simulation; $\lambda = \lambda' = 4.0$, $r =$ $r' = 16$ | 111 |
| 11. | Correlogram from Simulation; $\lambda = \lambda' = 2.0$, $r =$ $r' = 20$ | 112 |
| 12. | Correlogram from Simulation; $\lambda = 5.0$, $\lambda' = 5.0$, $r = 30$, $r' = 16$ | 113 |
| 13. | Correlogram from Simulation; $\lambda = 4.0$, $\lambda' = 4.0$, $r = 64$, $r' = 16$ | 114 |
| 14. | Correlogram from Simulation; $\lambda = \lambda' = 10.0$, $r = 50$, $r' = 30$ | 115 |
| 15. | Correlogram from Simulation; $\lambda = \lambda' = 5.0$, $r =$ $r' = 50$ | 116 |
| 16. | Correlogram from Simulation; $\mu = \mu' = 5.0$, $\sigma^2 =$ $\sigma'^2 = 6.0$ | 117 |

| | | |
|-----|---|-----|
| 17. | Correlogram from Simulation; $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 6.0$ | 118 |
| 18. | Correlogram from Simulation; $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 4.0$ | 119 |
| 19. | Correlogram from Simulation; $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 2.0$ | 120 |
| 20. | Correlogram from Simulation; $\mu = 4.0$, $\mu' = 20.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 4.0$ | 121 |
| 21. | Correlogram from Simulation; $\mu = 30.0$, $\mu' = 5.0$, $\sigma^2 = \sigma'^2 = 4.0$ | 122 |
| 22. | Correlogram from Simulation; $\mu = \mu' = 10.0$, $\sigma^2 = 2.0$, $\sigma'^2 = 0.0$ | 123 |
| 23. | Correlogram from Simulation; $\mu = 5.0$, $\mu' = 30.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 5.0$ | 124 |
| 24. | Correlogram from Simulation; $\mu = 5.0$, $\mu' = 30.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 5.0$ | 125 |
| 25. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 11 | 151 |
| 26. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 12 | 152 |
| 27. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 13 | 153 |
| 28. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 14 | 154 |
| 29. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 15 | 155 |
| 30. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 16 | 156 |
| 31. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 21 | 157 |
| 32. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 22 | 158 |
| 33. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 23 | 159 |

| | | |
|-----|--|---------|
| 34. | The Ratios $Q(k)$ and the Function ρ_k for Study No. 24 | 160 |
| 35. | Results from Simulation Study of Multiple Random Starts, Study No. 14 | 162-165 |
| 36. | Complex Activity Structure | 170 |
| 37. | Histograms for $Q(k)$ from Simulations with Gamma Distributed Span Lengths, $\lambda = \lambda' = 5.0$, and $r = r' = 50$ (Given Are Initial and Final Points with Interval Width in Parentheses) | 243 |
| 38. | Histograms for $Q(k)$ from Simulations with Gamma Distributed Span Lengths, $\lambda = 3.0$, $\lambda' = 2.0$, $r = 9$, and $r' = 14$ (Given Are Initial and Final Points with Interval Width in Parentheses) | 244 |
| 39. | Histograms for $Q(k)$ from Simulations with Gamma Distributed Span Lengths, $\lambda = \lambda' = 4.0$, $r = 64$, $r' = 16$ (Given are Initial and Final Points with Interval Width in Parentheses) | 245 |
| 40. | Histograms for $Q(k)$ from Simulation with Truncated Normal Distribution of Span Length $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 6.0$ (Given are Initial and Final Points with Interval Width in Parentheses) | 246 |
| 41. | Histograms for $Q(k)$ from Simulation with Truncated Normal Distribution of Span Length $\mu = 4.0$, $\mu' = 20.0$, $\sigma^2 = \sigma'^2 = 0.25$ | 247 |
| 42. | Histograms of Work and Idle Time | 249 |

SUMMARY

The general objective of this research is (1) to develop a cohesive theory regarding the application of systematic sampling techniques to activity sampling, where activity is classified into one of two states, and (2) to test this theory on several natural populations of activity. In order to attain the first part of the general objective, the following specific objectives were attained: (1) the development of formulas for the variance of the estimate of the population parameter, p , which is the fraction of the time period $[0, T]$ covered by the activity of interest, and the development of estimation procedures for the variance or an upper bound to the variance of the estimate, and (2) the development and test of measures for the classification of activity structure.

Two special cases and the general case were considered for the activity structure of simple activity, where simple activity is defined as the activity of single subjects, and complex activity is defined as the activity of groups of subjects.

In the first special case, the situation was considered in which the activity of interest occurs at most once on each time period $[0, T]$. In this case, systematic sampling with one random start on the interval $[0, d]$ and n total observations yields an estimate of the fraction of time that the activity is in the state of interest that is uniformly (in d) more precise than the corresponding estimate obtained from simple

random sampling. In 1948, Yates (164) outlined some mathematical methods for treating this situation, and his methods were employed herein. Yates' results were extended to consider the distribution of the estimate and a confidence statement on the parameter being estimated. Another extension was the utilization of stratification techniques where the study extends over several time periods such as shifts or half-shifts. There are no situations where systematic sampling should not be employed if the activity of interest occurs at most once per time period. An upper bound to the variance of the estimate is available for estimation of precision. Some examples of activities conforming to the assumptions of this special case are:

1. The routine maintenance and adjustment of a machine which occurs once during each time period such as a shift or half-shift.
2. The replenishment of raw stock, where the operator stops his usual work once during each period to replenish his stock.
3. The familiar coffee break, which occurs once per half-shift, whose length may nominally be 15 minutes, but in fact is a random variable.

In the second special case, the situation was considered in which the interval between observations is shorter than all occurrences of A and A' , where A is the activity of interest and A' is "not A ." This case was presented in 1960 by Davidson, Hines, and Newberry (47); however, the assumptions of the original presentation were not altogether correct, and these are investigated herein. Two subcases are

considered based on the number of "cycle length" variables added to form a given time to ℓ^{th} change in activity variable. The central limit theorem was employed for large ℓ and direct analysis was employed for the case of small ℓ with left truncated exponential densities of span lengths of occurrences of A and A' . Both considerations led to the statement of an easily calculated upper bound to the variance of the estimate of the parameter. Stratification techniques were again employed to extend the results to the case where sampling is conducted over several time periods, and the parameter of interest is the fraction of time for the entire period in which the activity is in the state of interest. The estimate, an upper bound to its variance, and confidence statements on the parameter are given for one period and multiple periods. Due to the unknown form of the distribution of the estimate, nonparametric techniques are employed in the statement of confidence. Systematic sampling is the superior design for most cases of this type encountered in practice, and this development is particularly applicable to large sample sizes and the utilization of photographic devices for the collection of data. Examples of activity which may permit a modest sample size due to the longer spans of A and A' are: (1) a repetitive activity cycle with "long" elements, (2) the operation of a process where equipment must be cleaned between the processing of batches, and (3) the utilization of commercial aircraft or other long cycle equipment utilization.

In the general case, where the assumptions of the two special cases may not be employed, the distributions of span lengths for A

and A' were considered in order to investigate the behavior of the serial correlation function. A covariance stationary, stochastic process was considered, and Cochran's important theorem dealing with systematic sampling (38, p. 117) was employed in showing that exponential distributions of these span lengths give rise to a monotone decreasing correlation function having sufficient properties to assure the superiority of systematic sampling over both simple random and stratified random sampling designs. Monte Carlo simulation methods were employed in the investigation of the correlation function when the span lengths have gamma and truncated normal densities. In these cases, several sets of parameters were used for each density, and the parameters were selected to give rise to several different values for the coefficient of variation in cycle length as well as different relative mean lengths for occurrences of A and A' . The correlation functions are not monotone decreasing, convex; however, further analysis was made, and the results indicate that systematic sampling is superior to simple random sampling when the interval between observations is not near an integral multiple of the mean cycle length. From the standpoint of the practitioner, if the interval between observations is less than three-fourths of the mean cycle length, and the ratio, $C = [\text{mean of the longer span length } (A, A')] / [\text{mean of the shorter span length } (A, A')]$, is less than three, then systematic sampling designs are superior or give rise to estimates having less variation when the coefficient of variation of the cycle length is greater than one tenth. In all cases, including extreme cases,

systematic sampling outperforms the corresponding simple random design when the interval between observations is less than one-fourth of the mean cycle length. An estimate of a "safe" upper limit for the interval between observations is developed, so that any interval greater than this value will provide a satisfactory estimate. The upper limit is a function of C and the coefficient of variation in cycle length. Gamma and truncated normal densities for the span lengths were selected because of practical considerations and the results of previous research on the form of the distribution of work elements.

Estimation procedures for estimating the variance of the estimate of the parameter were developed, and the two major cases considered were for one random start and multiple random start sampling designs. Confidence statements were developed for both designs, and stratification techniques were employed to extend these results to cover the case where sampling extends over several time periods such as shifts or half-shifts. Monte Carlo simulation was employed to validate the confidence statement for the multiple random start design.

The methods and formulas developed herein for simple activity structure were tested on ten natural populations of human activity, and systematic sampling performed well in each case. The results of these studies are presented graphically, and in some cases, the equivalent (in standard error) simple random sample would require roughly seven times as many observations as the systematic sampling design.

A brief investigation of complex activity indicated the applicability of estimation procedures developed in connection with simple

activity. Only minor modification was required to consider the variable measured as a fraction rather than a zero-one variable; however, the scope of this study did not include an extensive study of complex activity.

P A R T I

CHAPTER I

INTRODUCTION

Objectives of Study

The general objective of this research is (1) to develop a cohesive theory regarding the application of systematic sampling techniques to activity sampling of two state activities, and (2) to test this theory on several natural populations of activity. Systematic sampling is defined for both finite and infinite populations of "instants" of time. For finite populations, the instants each span constant width intervals of time which are numbered 1, 2, 3, ..., nk consecutively; the variable associated with the j^{th} interval is denoted by X_j , and a particular value (realization) of X_j is denoted by x_j . A systematic sample is selected by drawing an instant at random from the instants 1, 2, ..., k , and then selecting every k^{th} consecutive instant thereafter. For infinite populations representing a continuous realization of a process over the time period $[0, T]$, the variable associated with instant t is denoted by $X(t)$; a particular value (realization) of $X(t)$ is denoted by $x(t)$, and a systematic sample is selected by randomly drawing an instant on the interval $[0, d]$ and selecting the remaining observations at times $d, 2d, 3d, \dots, (n-1)d$ distant from the first observation. It is noted that $nd = T$.

Activity sampling is defined as the sampling of the activity of animate or inanimate subjects at preselected instants with the state of activity classified on each observation, where classifications are mutually exclusive and exhaustive.

The attainment of the first part of the general objective of this study is dependent upon the following two specific objectives: (1) the development of formulas for the variance of the estimate of the population parameter which is the fraction of the time period covered by the activity of interest, and the development of estimation procedures for the variance or an upper bound to the variance of the estimate, and (2) the development and test of measures for the classification of activity structure.

The Importance of the Study of Systematic Activity Sampling

The industrial engineer is the central figure in the industrial community who is concerned with the effective utilization of human, equipment, and material resources. The term "effective utilization" is used to imply a utilization that satisfies a criterion established by management. This criterion may be, for example, minimization of idle time, minimization of down time, minimization of secondary work time, and/or cost minimization. This research is not concerned with the appropriateness of the criterion.

The estimation of the proportion of time in which an activity is in a particular state has served in the past as a means for establishing and maintaining an effectiveness measure criterion. The two sampling techniques commonly employed for this estimation do not have standard nomenclature; however the terms "all-day time study" or "production study" and "activity sampling" are descriptive, and they will be employed herein. In usual practice, the all-day time study covers a relatively short time period from one to five days; the selection of a period is claimed to be either random or representative, meaning typical or average in some sense,

and instants are continuously observed. In usual practice, activity sampling study covers a longer period, and instantaneous observations are made at pre-selected instants. Various authors have claimed that activity sampling saves from fifty to eighty per cent of the cost of an equivalent production study; however this estimate of savings seems somewhat problematical, and in this claim equivalent production study is never defined.

The superiority of activity sampling for this estimation procedure is not questioned or argued here; however it is noted that the activity sampling study will cost less only if the observer can observe more subjects and/or profitably utilize his time between observations. The extent to which the observer's frequently interrupted alternate work time may be effectively used depends upon the alternate work available, the particular sampling plan employed, and the individual observer.

Since systematic sampling has the advantage of permitting the observer to establish an alternate work pattern as well as a pattern for observation times, recent literature indicates that several firms are now employing such sampling schemes. The bias* and precision** of the estimates have not been considered in many of these studies, and no complete theory has been available.

If systematic sampling plans are to be employed, there exists a need for a means of dealing with sampling error, and this development is essentially the responsibility of the industrial engineer who should bring

* The term "bias" relates to the centering of the probability distribution of an estimate $\hat{\theta}$, of the parameter θ . $\text{Bias} = E(\hat{\theta} - \theta)$.

** The term "precision" relates to the standard deviation of $\hat{\theta}$, and high precision implies a small value for the standard deviation.

theory and practice together, suggest improvements in practice, and solve problems in practice.

Definitions

The sampling design means the combination of a method for classifying the $N = nk$ population instants into k classes and a method of selecting one of the k classes, each having a designated probability of being selected. Employing the infinite population concept, both N and k are infinite, and the probability of selecting any single element or class of elements is zero. In this case the sample design means the method of selecting the n instants in the sample.

The sampling procedure is the operation of selecting one of the k classes by the method stated in the design, and the sample is the class obtained by the sampling procedure.

An unrestricted random sampling design to select n instants from a finite population of N instants is one in which there are $\binom{N}{n}$ classes, each having probability $1/\binom{N}{n}$ of being selected. One associated sampling procedure might consist of assigning an identification number to each class and selecting a number from a random number table to pick a specific class. In practice, the statistician would not employ such a procedure in most cases, since the classification is prohibitively expensive. Theoretically, this definition applies to the infinite population concept if N is allowed to approach infinity; however in practice, all infinite populations are observed as finite populations with N large.

Random sampling designs that are not unrestricted are said to be restricted, and systematic designs are one such restricted random design.

Stratified, cluster, and multiple stage designs represent some other restricted random designs.

A systematic sampling design is one in which the N instants are classified into k classes, S_1, S_2, \dots, S_k , where S_i consists of the elements $(i), (i+k), (i+2k), \dots, [i+(n-1)k]$, and a random sampling procedure is employed for selecting one of the S_i .

Scope of Study

This research is concerned with the development of measures useful in classifying the structure of activity durations as well as the determination of situations in which systematic sampling may be expected to be more precise or as precise as simple random sampling. There are to be few economic considerations per se, and systematic sampling will be considered superior to unrestricted random sampling if the standard deviation of the estimate of the parameter does not greatly exceed the standard deviation of the estimate calculated from an unrestricted random sample of the same size.

Sample data should be preserved in such a manner that the between-period variance component may be estimated, and this preservation is recommended for all studies regardless of the sampling design. This research, however, is concerned with the within-period variability and confidence statements rather than prediction or forecast intervals.

Simple activity is defined as the activity of single subjects, and complex activity is defined as the activity of groups of subjects. The major portion of this research, presented in Part II, is concerned with simple activity; however, complex activity is briefly treated in Part III.

A two way classification of activity is employed in the analytical development, and these activity states are denoted as A and A' (not A). Although there may be numerous classifications when the data are taken, the two way classification is employed in the statistical analysis where state A is the state of interest.

Study Procedure

Analytical and simulation methods are employed in this investigation, and several sets of data from natural populations of activity in the industrial community are used to form populations for testing the theory developed herein. Analytical techniques are employed where possible in the theoretical development; however, in the situations where analytical methods are either non-existent or inadequate, Monte Carlo simulation techniques are employed utilizing a Burroughs 220 Computer system.

A three step procedure is utilized in the study of simple activities. The first step is the description of the population with quantitative measures relative to the structure of activity duration. The second step consists of the development of systematic sampling designs and formulas for the variance or an upper bound to the variance of estimates calculated from samples. Methods for estimating the variance are developed where possible. The third step tests the methods and formulas developed in step two.

Assumption

It is assumed throughout this study that the activity being observed is insensitive to the observer or observation device. The activity state does not change due to observation. This assumption is employed in all

activity sampling studies; however when human subjects are observed, there may be a more marked departure from the condition stated in the assumption when systematic sampling designs are employed in lieu of unrestricted random designs or other restricted random designs. If the subjects are performing the activity in a large, open area where the approach of the observer is easily detected, all sampling designs lead to equally severe violations, and if observation can be made without the knowledge of the human subject, all designs completely satisfy the assumption; however there are many situations between these extremes where the assumption is more seriously violated for systematic sampling. Systematic sampling is not recommended for these cases, which are numerous in practice.

CHAPTER II

PRESENT STATE OF SYSTEMATIC ACTIVITY SAMPLING

Introduction

The objective of this chapter is to briefly review the significant literature of activity sampling with emphasis on the development of the theory and practice of systematic activity sampling. A more extensive review of the activity sampling literature is presented in Appendix A.

In an ideal situation, the practice of sampling is a composite of statistical theory, knowledge of the population being sampled, and experience gained through the use of various sampling designs and procedures. The history of sampling is difficult to trace as it stems from numerous roots with applications branching into government administration, education, agriculture, manufacturing, commerce, and several fields of science and engineering (145, p. 12). The argument most easily supported in regard to the development of the various sampling designs employed today is that these practices evolved gradually from simple notions. The earliest examples of sampling designs and procedures are found in such actions as taking a small portion of food for tasting and testing to determine the characteristics of the whole. The efforts of scientists to draw conclusions in regard to the laws of nature from observations in their environment is a sampling process, and Stephan (145) noted that: "all empirical knowledge is, in a fundamental sense, derived from incomplete or imperfect observation and is, therefore, a sampling of experience" (145, p. 3). In 1693, Halley's selection of data in Breslau to form the life table from which he drew

conclusions about the "mortality of mankind" (66, p. 596) was one of the first reported sampling studies in which the author recognized some of the implications of conclusions drawn from a sample.

The need for improved sampling methods developed gradually as the various bodies of knowledge expanded, and more sophisticated experimentation was undertaken. The development of unrestricted and restricted random designs has somewhat paralleled the need for such techniques. Stephan noted that:

In spite of successful experience, the use of random and systematic sampling procedures in statistical work made slow progress. At the Rome Session of the International Statistical Institute in 1925 a resolution was adopted recommending the use of sampling for statistical purposes with appropriate precautions as to its representativeness, mathematical statement of the precision, and full description of the methods employed. The reports that were submitted by the commission that drafted the resolution presented evidence of the usefulness of sound sampling procedures but both the reports and the resolution had less immediate effect on statistical practices than might have been expected

In America, serious attention was given to problems of sampling methodology by several committees of the Social Science Research Council ...

Two very important developments affecting the use of sampling in the United States occurred in 1933. One was organization of large scale work projects for the unemployed under the national programs of the Federal Emergency Relief Administration (1933-35), Civil Works Administration (1934-35), and Work Projects Administration (1935-40). The second was the enlistment of many leading statisticians from the universities and business in the reorganization of government statistical work and administration of emergency agencies (145, pp. 23-24)

The Development of Activity Sampling

The Initial Work

It was against the background described in the previous section that the initial work in activity sampling was accomplished and reported in 1935 by L. H. C. Tippett, an applied statistician employed in the British textile industry. Tippett's paper (148) reported the application of sampling

to the estimation of the fraction of non-productive time of machines and operators in a textile mill. Observers took periodic observation trips through the mills in question, recording the state of machines and operators as either idle or working. These rounds were made "about every hour" (148, p. 61); however there was some variability in the interval between rounds, and sampling error was described by the standard deviation associated with a binomially distributed proportion as follows:

$$\text{Standard error of } \hat{p} = \sqrt{\frac{p(1-p)}{n}}, \quad (1)$$

where \hat{p} is the estimate of the proportion of time, p , that the machine or operator was idle. The symbol n represents the number of observations. In the cases where the operator activity was independent and/or the machine activity was independent, the observations were pooled for operators and/or for machines in order to estimate machine idle time and operator idle time for the mill.

Observation tours were made for a number of days, and stratification effects were not considered; however Tippett recognized the presence of a variance component due to the day-to-day variability, and in the studies where a Pearson chi square "goodness of fit" test indicated that results obtained over several days did not conform to a binomial model with constant p , he concluded that this was due to the presence of day to day variability. Another possible explanation for some of this deviation is the presence of serial correlation in the activity and the non-randomness of observation times. The observation times in Tippett's studies were neither random nor systematic; some procedure between these extremes was utilized. Tippett's paper is reviewed in detail in Appendix A.

Since the time of this first paper, there have been more than one hundred and fifty publications reporting applications, criticisms, testimonials, and a few new developments in respect to the theory and methods of activity sampling. Most of these presentations have appeared in trade periodicals directed toward lower level management personnel; however, some which were presented in the professional journals, represent scholarly criticisms of the bases on which Tippett's method was formulated and/or a development and application of new theory.

Process Stability

Abruzzi (3, 4, 5) correctly noted an unstated but implied assumption underlying Tippett's work as the assumption of stability of the process being observed in order for the binomial model to be applicable. Abruzzi failed to recognize that a stable process will give rise to a day-to-day component of variance due to the process variability and the nesting of observations within days. This day-to-day component of variation should be employed when projecting the result of past observation into the future if a prediction interval is desired. In many of the papers following Tippett's article, much confusion existed about the population being observed and the differences between statistical estimation and prediction. Malcolm and Sammett (98) attempted to resolve some of this confusion for practitioners, and their paper is quoted extensively in Appendix A.

Other Sources of Variation

Cote and Scott (44) were the first researchers to develop and test a model to account for variation in prediction other than that of the sampling error associated with the observation of a single realization of a two valued process. Their paper presented a comparison between an all

day time study and a concurrent activity sampling study. The model equation was given for both studies as:

$$p_{ij} = \mu + w_i + d_j + (wd)_{ij} , \quad (2)$$

where p_{ij} is the proportion of time spent by worker i on day j on the task of interest, μ is the average over workers and days of the proportion of time spent on that task, w_i is the average over days by which worker i deviates from μ , d_j is the average over workers by which day j deviates from μ , and $(wd)_{ij}$ is the residual variation or worker i 's residual variation on day j (44, p. 32).

The sampling plan, for both studies, consisted of selecting "a" men from a larger group, and observing k of them each day until each man was observed b days. In the case of the activity sampling study, each man was observed at n random times per day. The estimates of μ are \hat{p}_c and \hat{p}_r respectively for the continuous and sampling study, and these are given as:

$$\hat{p}_c = \frac{1}{ab} \sum_i \sum_j \hat{p}_{ijc} , \quad (3)$$

and

$$\hat{p}_r = \frac{1}{ab} \sum_i \sum_j \hat{p}_{ijr} ,$$

where \hat{p}_{ijc} is calculated from the all day time study by dividing the total time the activity was in the state of interest by the total time of study for worker i and day j , and \hat{p}_{ijr} is calculated from the sampling study for worker i and day j by dividing the number of observations finding the activity in the state of interest by the total number

of observations, n .

The variances for the estimates are:

$$\sigma_{\hat{p}_c}^2 = \frac{1}{a} \sigma_w^2 + \frac{k}{ab} \sigma_d^2 + \frac{1}{ab} \sigma_{wd}^2, \quad (4)$$

and

$$\sigma_{\hat{p}_r}^2 = \frac{1}{a} \sigma_w^2 + \frac{k}{ab} \sigma_d^2 + \frac{1}{ab} \sigma_{wd}^2 - \frac{1}{nab} [(\sigma_w^2 + \sigma_d^2 + \sigma_{wd}^2) - \mu(1-\mu)]. \quad (5)$$

The estimation of σ_w^2 , σ_d^2 , and σ_{wd}^2 was accomplished by the technique of variance components analysis, and the results indicate that the sampling variance, shown in Equation (5) as the negation of the bracketed term with its multiplier, was small as compared to the other variance terms (44, p. 33).

The expression for the sampling variance in this study may be rewritten as:

$$\text{Within "nest" sampling error} = \frac{1}{nab} [\mu(1-\mu) - (\sigma_w^2 + \sigma_d^2 + \sigma_{wd}^2)]; \quad (6)$$

and when this expression is compared to the square of Equation (1), it is noted that the results are similar, with Equation (6) giving results less in magnitude due to the removal of process variability. The formula given by Tippett in 1935 and restated in Equation (1) of this thesis was employed in all reported applications between 1935 and 1956, the date of this paper, in the studies where sampling error was considered. There has been no widespread change in the industrial community since the time of this paper, and the literature indicates that most firms fail to distinguish between confidence type statements which are relative to the period sampled, and prediction intervals constructed on projections of estimates into some future

time period. None of the authors of the three texts on the subject of activity sampling has recognized that a distinction between estimation and prediction exists (22, 68, 62).

Groups of Subjects with Correlated Activities

Moder and Halladay (107) were evidently the first researchers to deal with group activity and correlation of activity within the group. Prior to the publication of their paper, studies were reported in which the correlation of activity within groups was ignored, and observations were erroneously treated as independent with the results pooled for the group. Several of these cases are reported in Appendix A (63, 108, 120). The groups studied by Moder and Halladay were crews of aircraft fabricators; and although there was correlation within the group, the activity between groups was independent. The variance for an estimate based on n observations of a crew with the activity of each crew member classified was given as:

$$\sigma^2 = \frac{p(1-p)}{nK}, \quad (7)$$

where K is the average number of "effective" observations per crew observation. A simulated study from file data of a continuous time study spanning one cycle consisting of twenty four consecutive shifts of this operation permitted the estimation of K by the following formula:

$$K = \frac{p(1-p)/n}{\sum_{i=1}^N (\hat{p}_i - \hat{p})^2 / (N-1)}. \quad (8)$$

The number of simulation "trials" was $N = 50$, \hat{p}_i is the estimate from the i^{th} trial, and \hat{p} is the average of the \hat{p}_i over the fifty trials.

It was found that the correlation of activity within the group was high, while correlation between groups was small. This approach is primarily of academic interest, since complete knowledge of the population is required. The paper was intended primarily to bring to light the error of assuming independence of observations when observing workers in groups.

Other Models

Conway (42) considered three basic models for possible application in the description of within period sampling error for various situations. They were: a binomial model for simple random sampling, a model for stratified sampling, and a model for non-independence and cluster sampling. In the case of non-independence, the application of cluster sampling techniques which consider the observation instants to be clusters of activity in the form of an m dimensional vector, $(x_{i1}, x_{i2}, \dots, x_{im})$, permits estimation of the variance of the estimate \hat{p} without knowledge of the K factor investigated by Moder and Halladay. However, in this study, m is assumed to be constant in time. The measurement associated with the i^{th} observation is:

$$y_i = \sum_{j=1}^m x_{ij} , \quad (9)$$

where $x_{ij} = 1, 0$, depending on whether the activity of the j^{th} member of the group is or is not in the class of interest at the time of the i^{th} observation. The estimate of p , the proportion of the time of all members in which the activity is in the class of interest, is given by \hat{p} , where

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n (y_i/m) . \quad (10)$$

The estimated variance of \hat{p} is given by $\hat{\sigma}_{\hat{p}}^2$, where

$$\hat{\sigma}_{\hat{p}}^2 = \frac{1}{n^2} \sum_{i=1}^n (y_i/m - \hat{p})^2. \quad (11)$$

In the case of stratified sampling of the activity of single subjects, Conway divides the total length of study into L strata, and f_h denotes the fraction of the total length of study in the h^{th} stratum. The symbols p_h , \hat{p}_h , p , and \hat{p} denote respectively the proportion of time and the estimate of the proportion of time in the h^{th} stratum and the total study period that the activity is in the class of interest. An unbiased estimate of p is given as:

$$\hat{p} = \sum_{h=1}^L f_h \hat{p}_h, \quad (12)$$

where

$$\hat{p}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi},$$

and $x_{hi} = 1, 0$ depending on whether or not the i^{th} random observation in the h^{th} stratum finds the activity in the class of interest. The variance of \hat{p} is given by $\sigma_{\hat{p}}^2$, where

$$\sigma_{\hat{p}}^2 = \sum_{h=1}^L \frac{(f_h)^2}{n_h} (p_h)(1-p_h), \quad (13)$$

and estimation of this quantity is not treated.

A group of three papers, presenting applications of survey sampling methods and theory to activity sampling, appeared in the years 1958, 1959,

and 1960. In the first of these, Rosander, Guterman, and McKeon (136) presented the methods employed by the Internal Revenue Service in the conduct of activity sampling. The design was stratified random with cluster sampling employed within the strata. The formulae for estimating p and estimating the variance of the estimate of p are presented (136, pp. 192-197). Mindlin (106) presented the methods employed by the Social Security Administration in conducting activity sampling studies. The government agencies, in this and the previous paper, employ the results from activity sampling studies in the preparation of budgets. Mindlin described a sophisticated, stratified, two stage plan with cluster sampling in the second stage. The design described is employed for the entire organization of 600 district offices ranging in size from five to eighty employees each. The complex estimation formulae are developed and sample calculations are given (106, pp. 288-294).

Halsey (67) applied the method of ratio estimation, a well known technique in the field of survey sampling, to the problem of worker activity analysis in a plant covering a large area where there was considerable worker mobility. Estimation formulae and sample calculations are presented (67, pp. 505-508).

The papers not reviewed in this section are briefly reviewed in Appendix A. These presentations were not particularly significant from the standpoint of adding to the basic knowledge of activity sampling; however they serve to document the applicability of sampling techniques to activity analysis, and they have been helpful to practitioners in that respect.

Systematic Activity Sampling

Eleven articles dealing specifically with systematic activity sampling were found in the literature. Of these, four have shown some attempt on the part of the author(s) to treat the problem of sampling error. The remaining papers have reported case studies.

In 1954, Owens (122) reported a theoretical development in which the proportion of time utilized by a secretary on telephone calls was the population parameter of interest. This activity is defined as an interrupter activity having the property that the time utilized for one occurrence of the activity is a random variable from occurrence to occurrence, and the occurrences were randomly located on the work day. Formulae for estimating the parameter and the variance of the estimate are developed (122, pp. 25-27); however there are some errors in the variance formula.

Kinniburgh and McTaggart (85) compared the results obtained from several concurrent systematic samples, random samples, and all-day time studies. In the case of systematic sampling or fixed interval sampling, they state:

In the fixed interval method the snap readings are made at fixed intervals of, say, one minute. A number of jobs have been studied this way, mainly bricklaying, excavating and site concreting. As a test of the method, continuous stop watch time studies were made on some of these at the same time, for comparison; in all cases there was good agreement between the results of the two methods (85, p. 180).

These authors were aware of the major difficulty associated with systematic sampling, and they warn others of employing a systematic sampling design with the interval between observations as a multiple of the period in cases where the occurrence of the activity of interest is periodic. The authors do not give formulae for the variance of the estimates, and they conclude:

The choice between random and fixed intervals must be left largely to personal preference in a particular situation. The random interval method is, of course, less affected by periodicity in work elements. On the other hand, the fixed interval method has many advantages. Not only is it simpler in operation, but for a given period of observation, it can be shown on statistical grounds that, on operations which are not periodic, it can give greater accuracy than the random interval method (85, pp. 81-82).

In 1955, Davis (49) authored the first paper in which there was a serious attempt to deal with the precision of systematic activity sampling. Three methods for selecting instants for observation are compared. They are:

1. instants occurring at regularly spaced time intervals (systematic sampling),
2. instants selected at random, with replacement, from a uniform distribution over a finite number of regularly spaced instants (finite population random sampling), and
3. instants selected at random from a uniform distribution over all possible instants in a given time interval (infinite population random sampling) (49, p. 111).

It is assumed that there is no observable trend in the mean to be estimated, and for the sake of convenience, all activity is classified into two states, A and A'. This is an article in which the parameter to be estimated is a process parameter rather than a proportion associated with a finite number of realizations of a process. A stationary stochastic process is assumed, and a realization of the process at time t is given by $x(t)$, where

$x(t) = 1$; for those instants, t , at which activity A is in progress.

$x(t) = 0$; for those instants, t , at which activity A' is in progress.

The stochastic process corresponding to $x(t)$ is denoted by $X(t)$, and it is assumed throughout Davis' work that

$$E[X(t)] = p, \text{ a constant.}$$

Observations are made at instants t_1, t_2, \dots, t_n on the time interval $[0, T]$. The estimate of p is denoted by $\hat{p}(n, T)$, where

$$\hat{p}(n, T) = \frac{1}{n} \sum_{i=1}^n X(t_i) . \quad (14)$$

It is noted that $E(\hat{p}) = p$, and that $\hat{p} \rightarrow p$ as $T \rightarrow \infty$ and $n \rightarrow \infty$, "except in the case where $X(t)$ is strictly periodic" (49, p. 112). The variance of the estimate is employed to describe its precision, and the conditional variance of \hat{p} for a fixed set of observation instants is given by the following relation:

$$\begin{aligned} \sigma^2(\hat{p} | t_1, t_2, \dots, t_n) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n P(t_i, t_j) - p^2 \\ &= \frac{p(1-p)}{n} \sum_{i=1}^n \sum_{j=1}^n \rho(t_i, t_j) , \end{aligned} \quad (15)$$

where $P(t_i, t_j)$ is the probability that both $X(t_i) = 1$ and $X(t_j) = 1$; and the relation $\rho(t_i, t_j)$, representing the correlation coefficient between $X(t_i)$ and $X(t_j)$, is given by:

$$\rho(t_i, t_j) = \frac{P(t_i, t_j) - p^2}{p(1-p)} . \quad (16)$$

The expected value of $\sigma^2(\hat{p} | t_1, t_2, \dots, t_n)$ is denoted by $\sigma^2(\hat{p})$, and the expressions for this expected variance are developed for each method of sample selection (49, pp. 113-114). The results for each of the three methods of sample selection listed previously are given as follows:

$$\text{Method 1: } \sigma^2(\hat{p}) = \frac{p(1-p)}{n^2} \sum_{i=1}^n \sum_{j=1}^n \rho(T_i, T_j), \quad (17)$$

where $T_{i+1} = T_i + d$, d is the interval between observations,
 $0 \leq T_i \leq T$, and $T_1 = 0$.

$$\text{Method 2: } \sigma^2(\hat{p}) = (p)(1-p) \left[\frac{1}{n} + \frac{(1-n^{-1})}{M^2} \sum_{i=1}^M \sum_{j=1}^M \rho(T_i, T_j) \right], \quad (18)$$

where the elements of the population are T_1, T_2, \dots, T_M .

$$\text{Method 3: } \sigma^2(\hat{p}) = (p)(1-p) \left[\frac{1}{n} + \frac{(1-n^{-1})}{T^2} \int_0^T \int_0^T \rho(t, \tau) dt d\tau \right]. \quad (19)$$

An illustrative example was considered in which the lengths of time over which $X(t) = 1$ are a_j , and the lengths of time over which $X(t) = 0$ are denoted a'_j for the j^{th} occurrence of activity A and A' respectively. Both a_j and a'_j are mutually independent for $j = 1, 2, \dots$, and they have the following densities:

$$\begin{aligned} g_a(a_j) &= \lambda_a e^{-\lambda_a a_j} ; & a_j &\geq 0 \\ &= 0 & ; & a_j < 0 \end{aligned} \quad (20)$$

and

$$\begin{aligned} g'_a(a'_j) &= \lambda'_a e^{-\lambda'_a a'_j} ; & a'_j &\geq 0 \\ &= 0 & ; & a'_j < 0 . \end{aligned}$$

The mean span lengths are $1/\lambda_a$ and $1/\lambda'_a$ respectively for occurrences of A and A'. It is noted that

$$p = \frac{\lambda_{a'}}{\lambda_a + \lambda_{a'}} \quad (21)$$

and the function $\rho(t, \tau)$ is determined to be

$$\rho(t, \tau) = e^{-\frac{|t-\tau|}{t_0(p-p^2)}}, \quad (22)$$

where

$$t_0 = \frac{\lambda_a + \lambda_{a'}}{\lambda_a \lambda_{a'}}.$$

It is assumed that $T \gg t_0(p-p^2)$ and $n \gg 1$ in order to neglect terms in the second order of magnitude in $(1/T)$, and the variances associated with the three methods of sample selection are given as follows:

$$\text{Method 1: } \sigma^2(\hat{p}) \cong \left[\frac{(p-p^2)}{T} \right] \left[\frac{\delta(1+e^{-Y\delta})}{(1-e^{-Y\delta})} \right], \quad (23)$$

where

$$Y = [t_0(1-p)p]^{-1},$$

and

$$\delta = T/(n-1).$$

$$\text{Method 2: } \sigma^2(\hat{p}) \cong \left[\frac{(p-p^2)}{T} \right] \left[\frac{\delta(1+e^{-Y\delta})}{(1-e^{-Y\delta})} + \Delta \right], \quad (24)$$

where

$$\Delta = T/(n-1),$$

and

$$\delta = T/(M-1).$$

$$\text{Method 3: } \sigma^2(\hat{p}) \cong \left[\frac{(p-p^2)}{T} \right] \left[2(p-p^2)t_0 + \Delta \right], \quad (25)$$

where

$$\Delta = T/(n-1) .$$

In order to compare the results of Equations (23), (24), and (25) with the results obtainable from a timing of a finite interval $[0, T]$, Davis gives the estimate of p from continuous timing as p^* , where

$$p^* = \frac{1}{T} \int_0^T X(t) dt, \quad (26)$$

and the variance of this estimate is shown to be $\sigma^2(p^*)$, where

$$\sigma^2(p^*) = \frac{p(1-p)}{T^2} \int_0^T \int_0^T \rho(t, \tau) dt d\tau. \quad (27)$$

For the process presented in the preceding paragraph, this expression reduces to

$$\sigma^2(p^*) \cong (2t_0/T)(p-p^2)^2, \quad (28)$$

and numerical comparisons indicate that the systematic sampling design yields an estimate having less variance than either of the other proposed designs (49, p. 116). Of course the extent to which the variance is less depends on the particular process.

Davis points out that the correlation function gives the whole picture when comparing systematic to random sampling; however he fails to note the difficulties associated with the determination of this function when the process has some behavior other than the Markovian behavior associated with the densities of his illustrative example. It is noted that

a covariance stationary process with a non-increasing correlation function will always indicate the superiority of systematic over simple random sampling. In Davis' words, "Roughly speaking, p will have this property if there is no strong periodic aspect to the process" (49, p. 117).

In 1957, Conway (42) briefly noted that systematic sampling was gaining in popularity, and he stated: "If there is no cyclic behavior present in the phenomenon under study, systematic sampling is acceptable, and certainly advantageous from an operating point of view" (42, p. 111). In the following year, Haines (64) reported the results of an experiment in which several systematic sampling studies, employing a two minute interval between observations, were compared to a known value of p , the idle time proportion. Two sets of comparisons were made. In the first set, file data, consisting of a continuous time log of worker activity, was sampled. In the second set, concurrent continuous timing and fixed interval observation by independent observers furnished the data for comparison. The activity studied was that of human subjects, and the operation conducted by the subjects was of a servicing nature with no definite cycle. "The work requirements were both varied and restricted by the demand for service" (64, p. 267). Activity was classified as either working or idle, and the values compared were the fraction of idle time from the continuous log or concurrent continuous timing with the fraction of idle time estimate from the systematic sample data. In all cases good agreement was obtained (64, p. 267).

In 1959, Wolff (162) described a method for activity sampling employed in connection with physiological research. A one minute fixed interval was used as the interval between observations, and an observer

utilized a "walkie talkie" to announce the activity of twelve subjects in preassigned order. A recording machine was operated by a clerical employee who operated a receiver as well as the recording machine. The advantage of this procedure is claimed to be the reduction in man-hours required to observe a group of this size. There was no treatment of sampling error, and no numerical results are given.

In the following year, Davidson, Hines, and Newberry (47) developed a theoretical formula for the variance of the estimate obtained by systematic sampling when certain assumptions are satisfied. The assumptions were:

1. the interval between observations is less than the span of activity A, and A' (not A), for all occurrences of A and A',
2. the distribution of y_i is uniform on $[-d/2, d/2]$, where y_i is the error associated with the i^{th} change in activity, and d is the interval between observations (see Figure 1); and
3. the y_i are independent random variables for $i = 1, 2, \dots, N'$, where N' is the number of occurrences of A during the study period.

Under these assumptions, the estimate of the proportion, p , of the interval $[0, T]$ covered by activity A is \hat{p} where

$$\hat{p} = \frac{1}{n} \sum_{j=1}^n x_j, \quad (29)$$

and $x_j = 1, 0$, depending on whether or not the j^{th} systematic observation finds the activity in the class of interest. The standard deviation of \hat{p} was shown to be approximately $\hat{\sigma}_{\hat{p}}$, where

$$\hat{\sigma}_{\hat{p}} = \frac{1}{n} \sqrt{\frac{N'}{6}} \quad (30)$$

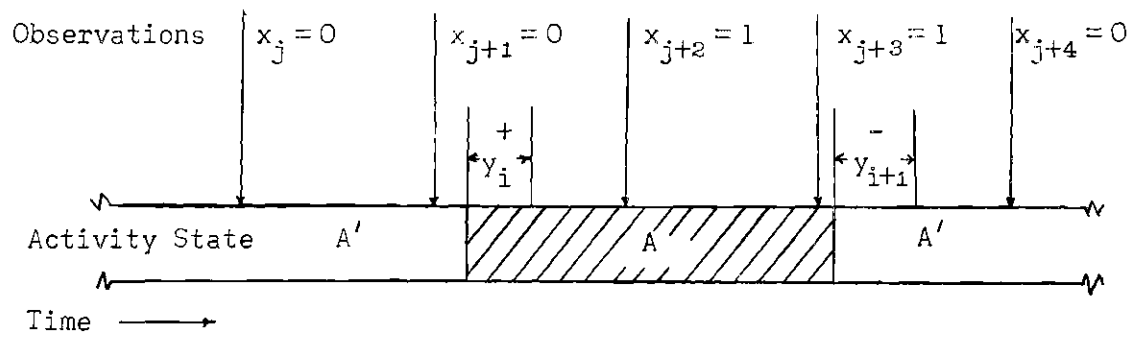


Figure 1. Error Associated with Systematic Sampling

The property of bias was not treated, and the authors were not aware of Davis' excellent paper (49).

In 1960, Isherwood (77) reported the use of a systematic sampling design for sampling the activity of employees in a small plant nursery. The interval between observations was thirty minutes, and the binomial variance relation was employed for the precision estimates. This relation may lead to either over-estimation or under-estimation in this case, and Isherwood did not consider this possibility; however, he referenced Davis' paper (49), and he justified the use of systematic sampling on the basis of Davis' statement that it is applicable when no strong periodic element is present.

Provost (128) presented the results of his research comparing "the use of memo-motion film for analyzing non-work activities with three other methods: a concurrent systematic interval study, a systematic sampling of the film, and a random sampling of the film" (128, p. 1). In comparison, the results obtained from systematic sampling in both comparison studies showed good agreement with the results determined from a frame by frame analysis of the film, and the estimate was better than the random sample estimate (151, pp. 24-42).

A recent paper, reporting an application of systematic sampling, was authored by Timmins (147) and published in November, 1961. The results of a systematic sampling study of the activity of three overhead cranes in a foundry are reported. There is no treatment of sampling error, and in the writer's opinion this paper is not a significant contribution to the literature.

The most recent paper dealing with systematic activity sampling was presented in July, 1963 by Flowerdew and Malin (64). Cycles of activity are considered, and the activity of interest occurs as an element of the cycle. Systematic sampling intervals are classified into three classes as:

- Type 1. Intervals that nearly coincide with the cycle time or some simple multiple or proportion of it.
- Type 2. Intervals different from the cycle time, but longer than the element time.
- Type 3. Intervals shorter than the element time (64, p. 202).

It is hypothesized that Type 1 gives an estimate with large variance, that the binomial variance formula is adequate for the treatment of sampling error associated with Type 2, and that the formula for estimation of variance given by Davidson, Hines and Newberry (47) is adequate for situations of Type 3.

The hypotheses were tested by computer simulation where the element durations were taken as truncated normal random variables, and the cycle has four elements. Although the number of simulations was inadequate for any strong inference, the hypotheses are reported to be accepted.

The experimental details were omitted which precludes an objective evaluation of the experiment and the conclusions. The authors were not specific in the statement of Type 1, Type 2, and Type 3 situations; so that any sampling interval qualifies for treatment under Type 1.

A recommended study procedure includes the preliminary timing of five cycles to estimate the mean cycle length.

Discussion of Results

Activity sampling has been employed for the analysis of numerous types of activity since the introduction of the technique in 1935; however

adequate formulae for the estimation of sampling error have lagged behind applications. There are four areas in which further research is needed. First, adequate designs and procedures for dealing with the between period variance component should be developed for utilization in the industrial community; second, the techniques successfully employed in the field of survey sampling should be investigated further for possible application to the activity sampling problem; third, techniques are needed for increasing the relative precision of the estimate, \hat{p} , of the parameter p , when p is small; and fourth, a theoretical and experimental research project to investigate the applicability of systematic sampling has not been presented prior to this thesis.

In so far as systematic activity sampling is concerned, the seven cited references in which applications have been reported indicate an increasing interest in systematic activity sampling, and in no cases were difficulties encountered by the practitioners. Davis' paper was by far the most significant contribution to the theory of systematic activity sampling. Owens' paper was of limited value, the paper by Davidson, Hines and Newberry was not as complete as it might have been; although it deals with a special case defined under limiting assumptions. The conclusions of Flowerdew and Malin are probably correct in part, although their argument is inadequate.

None of the authors investigating the applicability of systematic activity sampling referenced the existing literature of systematic sampling that has appeared in the various statistical journals, and the following chapter has the review of this literature as its objective.

CHAPTER III

A REVIEW OF THE SYSTEMATIC SAMPLING LITERATURE

Introduction

The purpose of this chapter is to review the literature of systematic sampling. Generally, the literature shows an interplay between practical needs and the provision of statistical techniques for their satisfaction. Buckland states:

Just as the methods of Factor Analysis grew up in the atmosphere of the psychologist, so has Systematic Sampling grown up alongside problems of forestry and land use (36,p.208).

In some of the engineering and physical sciences, forms of systematic sampling have been in use for many years, and other applications have been in agriculture, forestry, and meteorology. The definitions of systematic sampling must be generalized to include spatial as well as temporal variables in order to review this literature. There are two kinds of systematic sampling emerging from similar methods of sample selection; however the interpretation is different. In the first kind, the population elements are arranged in a random order, and systematic sampling is employed to obtain a random sample. This is frequently referred to as quasi-random sampling, and a number of papers dealing with this method are not included in this review.

Early Papers Relative to Systematic

Sampling Applications

In 1922, R. A. Fisher (57) reported the results of an extensive study of rainfall data in England. The objective of this research was to investigate the correlation of rainfall data at different weather stations. Data for forty years of rainfall, divided into weekly periods, supplied about 2000 data points, and the function

$$\rho_u = \tanh^{-1}(u^{-3/5})$$

was suggested for the correlation between weekly rainfall at two weather stations a distance of u apart. The importance of this work lies in the presentation of the correlation function or correlogram as a characteristic of an autocorrelated population.

Hasel (69) reviewed the pre-World War II literature of forestry sampling in 1938. He stated conditions under which the sample provides the necessary information for estimating sampling error, and he pointed out that sampling done by "cruising" at fixed intervals violates the conditions for valid estimation of sampling error. Auxiliary random observations were suggested as a possible means of estimating sampling error; however the practice was noted as an expensive one. Hasel suggested that systematic sampling be retained due to its advantages in administration and mapping; however he emphasized the need for adequate methods of dealing with sampling error.

In 1942, two papers were presented in which land usage was measured by means of systematic sampling. Proudfoot (127) applied the traverse surveying technique to the problem of estimating the adequacy

of sample data, and indices were calculated as functions of the correlations between the estimates at different spacings of the traverse lines. Osborne (121) presented what seems to be the first attempt to deal with the underlying position of serial correlation as playing the major role in systematic sampling. He suggested the function

$$\rho_u = e^{-\lambda u}$$

to describe the serial correlation in forestry and land usage surveys.

In comparing various methods for sampling tree nursery inventories, Johnson (78) concluded:

A systematic sample is to be recommended under certain conditions. It should be realized, however, that if a valid estimate of sampling error is demanded of the sample, a random method of selecting the sample must be used, for there is yet no theory for samplings other than random (78, p. 678).

The Initial Development of Theory

The following year, 1944, was an important year in the development of systematic sampling in that the first theoretical treatment was presented. The authors, W. G. and L. H. Madow (92, 93, 94, 95), have been prominent contributors throughout the development of the present theory of systematic sampling. This first paper (93) considered unstratified and stratified systematic sampling from a particular finite population of elements. The Madows defined \bar{x}_i as the mean of the elements of S_i , the i^{th} possible systematic sample of the every k^{th} type.

$$n\bar{x}_i = x_i + x_{i+k} + x_{i+2k} + \dots + x_{i+(n-1)k} \quad (31)$$

The symbol \bar{x} represents the sample mean, that is $\bar{x} = \bar{x}_i$ if S_i is selected by the sampling procedure. The following important theorem is proved:

Theorem: Using the systematic sampling design, the estimate \bar{x} is an unbiased estimate of μ' , and \bar{x} has variance given below, where μ' is the mean of the finite population.

$$\begin{aligned} \sigma_{\bar{x}}^2 &= \frac{\sigma^2}{n} \left[1 + (n-1) \bar{\rho}_k \right] \\ &= \frac{\sigma^2}{n} \left[1 + 2 \sum_j \rho_{kj} \right] \\ &= \frac{\sigma^2}{n} \left[1 + \frac{2}{n} \sum_{\delta} (n-\delta) \rho'_{k\delta} \right] \quad (93, p.6) . \end{aligned} \quad (32)$$

The population variance, σ^2 , is defined as follows:

$$\sigma^2 = \frac{\sum_{i=1}^{nk} (x_i - \mu')^2}{nk} , \quad (33)$$

and the quantities ρ_{kj} , $\rho'_{k\delta}$, and $\bar{\rho}_k$ are defined below.

$$1. \quad \rho_{kj} = \left[\frac{1}{\sigma^2} \frac{\sum_{v=1}^{nk} (x_v - \mu') (x_{v+kj} - \mu')}{nk} \right] , \quad (34)$$

where ρ_{kj} is the serial correlation between elements kj units apart, and if $h > nk$, then $x_h = x_{h-nk}$. This is frequently termed the circular definition of serial correlation.

$$2. \quad \rho'_{k\delta} = \frac{1}{\sigma^2} \left[\frac{\sum_{j=1}^{k(n-\delta)} (x_j - \mu') (x_{j+k\delta} - \mu')}{k(n-\delta)} \right], \quad (35)$$

where $\rho'_{k\delta}$ is the serial correlation between elements $k\delta$ units apart.

$$3. \quad \begin{aligned} \bar{p}_k &= \frac{2}{n-1} \sum_j \rho_{kj} \\ &= \frac{2}{n(n-1)} \sum_{\delta} (n-\delta) \rho'_{k\delta}, \end{aligned} \quad (36)$$

where \bar{p}_k is termed the intraclass correlation coefficient. Three basic results were stated as follows:

1. If the serial correlations have a positive sum, systematic sampling is worse than random sampling. 2. If the serial correlations have a sum that is approximately zero, systematic sampling is approximately equivalent to random sampling. 3. If the serial correlations have a negative sum, systematic sampling is better than random sampling (93, p.2).

Systematic sampling with multiple random starts is considered, and in the case in which g of the S_i are randomly selected without replacement, it is shown that

$$\bar{\bar{x}} = \frac{1}{g} \sum_{\beta=1}^g \bar{x}_{\beta} \quad (37)$$

is an unbiased estimate of μ' . The variance of $\bar{\bar{x}}$ is shown to be

$\sigma_{\bar{\bar{x}}}^2$, where

$$\sigma_{\bar{\bar{x}}}^2 = \frac{k-g}{k-1} \cdot \frac{1}{g} \cdot \sigma_{\bar{x}}^2, \quad (38)$$

and an unbiased estimate of this variance is shown to be $\hat{\sigma}_{\bar{x}}^2$, where

$$\hat{\sigma}_{\bar{x}}^2 = \frac{k - g}{k - 1} \cdot \frac{1}{g(g - 1)} \sum_{\beta=1}^g (\bar{x}_{\beta} - \bar{\bar{x}})^2. \quad (39)$$

Two stratified designs are considered for stratified systematic sampling. The first assumes that $k_1 = k_2 = \dots = k_L = k$, where there are L strata, and k_h is the interval between observations in the h^{th} stratum. The sampling procedure consists in selecting one of the integers $1, 2, 3, \dots, k$ at random with each integer having probability $1/k$ of being selected. If the integer i is selected, the sample obtained from the h^{th} stratum is $x_{hi}, x_{hi+k}, \dots, x_{hi+(n_h-1)k}$; for $h = 1, 2, 3, \dots, L$, and there are k possible samples. The second procedure consists in selecting one of the integers $1, 2, \dots, k_h$ at random for each stratum. In this case, there are $k_1 \cdot k_2 \cdot \dots \cdot k_L$ possible samples, and the probability of a particular sample being selected is $1/(k_1 \cdot k_2 \cdot \dots \cdot k_L)$. The estimate of the population mean μ' is the same for both procedures, and this estimate, given below, is unbiased.

$$\bar{x} = \sum_{h=1}^L \frac{N_h}{N} \bar{x}_h, \quad (40)$$

where \bar{x}_h is the sample mean in the h^{th} stratum, N is the population size, and N_h is the number of elements in the h^{th} stratum.

The variance of the estimate is given for the first and second procedure respectively below.

$$1. \quad \sigma_{\bar{x}}^2 = \frac{1}{N^2} \sum_a \sum_b N_a N_b \left[\frac{1}{k} \sum_i (\bar{x}_{ai} - \mu'_a) (\bar{x}_{bi} - \mu'_b) \right], \quad (41)$$

where μ'_a , μ'_b , \bar{x}_{ai} , and \bar{x}_{bi} are the stratum means and i^{th} possible systematic sample means in strata a and b respectively.

$$2. \quad \sigma_{\bar{x}}^2 = \frac{1}{N^2} \sum_{h=1}^L N_h^2 \sigma_{\bar{x}_h}^2, \quad (42)$$

where $\sigma_{\bar{x}_h}^2$ is the value shown in Equation (32) except the calculation is made within stratum h .

Three types of functions are presented to describe the population, and systematic sampling is compared in precision to simple random and stratified random sampling for each function; however no general results were obtained, and in the case of the periodic function investigated, the precision of systematic sampling was found to be closely related to the period of the function and the interval between observations.

A simplified version of this paper was presented by one of the authors (92, Madow, L. H.), and the approach in this second paper was more definitely related to cluster sampling with one cluster sampled and no subsampling within the cluster. In the case of unstratified populations, the estimate of the mean was the same as that stated in Equation (31), and the variance of the estimate, for a particular value of k , was given as follows:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} + \frac{(n-1)}{n} \left[\frac{1}{k} \sum_{i=1}^k (\bar{x}_i - \mu')^2 - \frac{1}{nk(n-1)} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right], \quad (43)$$

where \bar{x}_i is the mean of the i^{th} possible systematic sample, and x_{ij} is the j^{th} member of the i^{th} possible systematic sample.

The author concludes from experience gained in the application of systematic sampling that:

In the applied cases where a systematic design is more efficient than a stratified random design, the systematic design is about twice as efficient as the stratified random design, whereas in most of the cases in which the systematic design is less efficient than the stratified random design, the stratified design has only a slight gain over the systematic design (92, p. 217).

Increased Interest in Systematic Sampling

Before considering the next development in the theory, two papers dealing with practical problems will be reviewed. Deming and Simmons (50) presented a method used for sampling dealer tire inventories during the tire rationing period of World War II. This paper, published in 1946, described a method combining stratification of the population and a systematic design within strata. It seems that nonstatistical considerations were responsible for this design and the work of the Madows was not referenced. Homeyer and Black (75) employed several designs in estimating the yield of oat fields at the Iowa State Experimental Farm. Complete data were available so that there was a basis for comparison of sampling designs. The systematic design yielded a design slightly worse than the other designs; however, the authors state:

It (systematic sampling) required less time and can be used with less skilled labor than is required with the other methods. Data are not yet available to adjust the relative efficiencies for time needed to take the samples with alternate designs. Workers at the Iowa Experiment Station who have used all four methods believe that cutting strips with power mowers gives the most information per hour of sampling (75, p. 343).

The Correlogram in Analyzing Systematic Designs

A paper by Cochran (38) in 1946 related the form of the correlation function to the relative precision of systematic sampling. The paper represents a radical departure from the work of the Madows in that the parameters of interest are regarded as process parameters rather than characteristics associated with one realization of the process; the reader will recall that this approach was also used by Davis (49) in the important paper reviewed in Chapter II. In this situation, the population consists of $N = nk$ random variables: X_1, X_2, \dots, X_N ; and the case considered is the one in which

$$E(X_i) = \mu \quad , \quad (44)$$

$$E(X_i - \mu)^2 = \sigma^2 \quad ,$$

$$E(X_i - \mu) (X_{i+u} - \mu) = \rho_u \sigma^2 \quad ,$$

and

$$\rho_u > \rho_v \quad \text{if} \quad u < v \quad .$$

The symbols σ_r^2 , σ_{st}^2 , and σ_{sy}^2 are used to denote the average variances of the sample means of random, stratified random, and systematic designs. "This average being taken over all finite populations from the infinite population specified" (38, p. 167). According to the definitions in Chapter I, this would be considered a finite population having an infinite number of outcomes or realizations. Cochran shows in further development that:

$$\sigma_r^2 = \frac{\sigma^2}{n} \left(1 - \frac{1}{k}\right) \left[1 - \frac{2}{nk(nk - 1)} \sum_{u=1}^{nk-1} (nk - u) \rho_u\right] \quad , \quad (45)$$

$$\sigma_{st}^2 = \frac{\sigma^2}{n} \left(1 - \frac{1}{k}\right) \left[1 - \frac{2}{k(k-1)} \sum_{u=1}^{k-1} (k-u) \rho_u\right], \quad (46)$$

and

$$\sigma_{sy}^2 = \frac{\sigma^2}{n} \left(1 - \frac{1}{n}\right) \left[1 - \frac{2}{nk(k-1)} \sum_{u=1}^{nk-1} (nk-u) \rho_u + \frac{2k}{n(k-1)} \sum_{u=1}^{n-1} (n-u) \rho_{ku}\right]. \quad (47)$$

Equations (44) are assumed to be true, and the variances of Equations (45), (46), and (47) are compared, with the results stated as an important theorem in the theory of systematic sampling, which deals with populations having a convex, decreasing correlogram. The proof is given by Cochran (38, p. 117).

Theorem: For the populations in which $\rho_i \geq \rho_{i+1} \geq 0$, $i = 1, 2, \dots, nk-2$, and

$\delta_i^2 = \rho_{i-1} + \rho_{i+1} - 2\rho_i \geq 0$; $i = 1, 2, \dots, nk-2$, then

$$\sigma_{sy}^2 \leq \sigma_{st}^2 \leq \sigma_r^2,$$

for any size sample. Further $\sigma_{sy}^2 < \sigma_{st}^2$ unless $\delta_i^2 = 0$; $i = 1, 2, \dots, nk-2$ (38, p. 117).

Cochran states: "So far as practical applications are concerned, the restriction that ρ_u should be concave upwards may not be severe" (38, p. 173). He points out that the correlograms suggested by both Osborn and Fisher satisfy this requirement.

In 1948, Nordskog and Crump (119) reported an application of systematic sampling methods to poultry farming. The authors compared results obtained with egg sampling designs based on random days, consecutive days and interval days. Differences between the three methods

were found to be very small. In a paper reporting experimental evidence from several row crop surveys, Haynes (71) found stratified random sampling to be much more precise than a systematic grid sample selected by superimposing a grid over the area and selecting squares at fixed intervals; however a zig-zag systematic sample, selected at fixed intervals along a zig-zag line drawn across the area, compared favorably with stratified random sampling. Finney (55) empirically studied the relationship between systematic and random sampling in timber surveys. Data from two extensive studies involving complete enumeration of forest timber are employed in this comparison which is formed by a Monte Carlo sampling from the file data. The following conclusion is justified in this study.

Detailed comparisons reported earlier may be summarized by the statement that systematic sampling proved more precise than either form of stratification, but that the consequent narrowing of the limits of error is scarcely sufficient to justify a general recommendation that systematic sampling always be used (55, p. 97).

An important paper was presented by Yates (164) in 1948, and a summary of this paper was placed in perspective in the book, Sampling Methods for Census and Surveys (165), authored by Yates and published in 1949. This book has become a standard reference in England. Yates' paper in 1948 has been recognized as a landmark in the literature of systematic sampling for several reasons. First, it gives the reader an immediate sense of the practical aspects and utility of systematic sampling; second, this work is the first development in which attribute variables are considered; and third, end corrections are proposed for

the elimination of trend errors in systematic sampling. It is shown that it is impossible to make fully reliable estimates of sampling error from systematic sampling results alone, and methods for using supplementary observations are described. The performance of systematic sampling is investigated for the following population models: (1) a two-valued function, (2) normally distributed material, and (3) a one-term autoregressive function. Systematic sampling is shown to be uniformly more precise than random or stratified random sampling of a two-valued realization on which the activity of interest spans at most one interval in the period $[0, T]$. The one-term autoregressive model gives rise to an exponential correlation function, and this function clearly satisfies the conditions required for the application of Cochran's theorem.

Problems associated with estimation are discussed, and Yates suggests a division of the sample into nine parts for the estimation of sampling error. Only one particular set of data is considered, and the estimate suggested is $\hat{\sigma}_{sy}^2$, where

$$\hat{\sigma}_{sy}^2 = \frac{N-n}{Nn} \sum_{u=1}^g \frac{d_u^2}{7.5g}, \quad (48)$$

$$d_1 = \left(\frac{1}{2}x_1 + x_3 + x_5 + x_7 + \frac{1}{2}x_9 \right) - (x_2 + x_4 + x_6 + x_8),$$

$$d_2 = \left(\frac{1}{2}x_9 + x_{11} + x_{13} + x_{15} + \frac{1}{2}x_{17} \right) - (x_{10} + x_{12} + x_{14} + x_{16}),$$

...

g is the number of differences the sample provides, and N is the population size. Yates found this estimate to be an overestimate in most

applied cases where no strong periodic effect was present; however the estimate itself is based on much intuitive reasoning.

Jones (79) and Kendall (84) considered the problem of employing systematic sampling to estimate the average value of a random variable in a homogeneous, one-dimensional population that depends on one parameter. These papers, presented in 1948, consider an autoregressive population behavior function, and the central problem treated is that of distributing sample members according to a criterion relating to the degree of correlation between successive members or positions in the population. Kendall treated the problem of evaluating the relative merits of increasing the sampling intensity within a strip or extending the strip. The general equations presented for this comparison have not been solved. No new, definite results were obtained in either of these mathematical papers. Kendall concludes by adding his warning to that of others that systematic sampling should not be employed when there is a strong periodic element.

Two papers dealing with theoretical developments appeared in 1949. In the first of these, W. G. Madow (94) extended the results of his earlier work to include the consideration of the systematic sampling of clusters of elements. In the second paper, Quenoville (129) presented the major work in two-dimensional systematic sampling. Madow first treats stratified sampling, and he considers an estimate, y , as a linear combination of "contributions from the strata." The estimate, y , is given as

$$y = y_1 + y_2 + \dots + y_L, \quad (49)$$

where y_h is the contribution from the h^{th} stratum. It is pointed out that if the selection of elements from different strata is done in such a way that the contributions from different strata are negatively correlated, the variance of the estimate will be less than in the case in which contributions are independent but have the same covariances within strata. In this case,

$$C' = \sum_{i \neq j} \sigma_{y_i y_j} < 0, \quad (50)$$

where $\sigma_{y_i y_j}$ is the covariance between y_i and y_j . Sampling designs having this property are said to possess "negative correlation" (94, p. 336).

In the development of systematic cluster sampling, Madow follows Cochran's logic in considering a particular finite population as a single realization of the variables associated with the finite number of population elements. By considering the expected value of the value C' , defined in Equation (50), Madow derives a result similar to Cochran's theorem giving the sufficient conditions for systematic sampling to be more precise than random designs. For cluster sampling, the population consists of N clusters of elements, and each cluster has M elements, where $N = nk$, and $M = cm$. The random variable associated with the α^{th} element of the i^{th} cluster is denoted by $X_{i\alpha}$, and a particular value of $X_{i\alpha}$ is denoted by $x_{i\alpha}$. For a particular realization, the population mean is denoted by μ' , where

$$\mu' = \frac{1}{MN} \sum_{i=1}^N \sum_{\alpha=1}^M x_{i\alpha}. \quad (51)$$

In the situation in which there is complete enumeration of all the clusters in the sample, the sampling design considered requires the systematic selection of n of the N clusters. The sample mean is given by \bar{x} , where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n \bar{x}_i, \quad (52)$$

and \bar{x}_i is the mean of the i^{th} cluster in the sample. The variance of the mean is shown to be $\sigma_{\bar{x}}^2$, where

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{nM} [1 - (M - 1) \rho] [1 + (n - 1) \bar{\rho}_k], \quad (53)$$

and σ^2 is the total population variance; $\bar{\rho}_k$ is the expected value (over realizations) of the value defined in Equation (36), and ρ is the expected value of ρ^* , where

$$\rho^* = \frac{1}{\sigma^2} \left[\sigma_b^2 - \frac{\sigma_w^2}{M - 1} \right], \quad (54)$$

and σ_b^2 and σ_w^2 are the between systematic samples variance and the within systematic sample variance respectively. These quantities were defined by L. H. Madow in 1946 (92). No estimation procedures are suggested.

In the paper entitled, "Problems in Plane Sampling," Quenoville (129) extends Cochran's results, and he proceeds to treat problems in two-dimensional sampling. Combinations of random, stratified random, and systematic sampling designs are considered for the complete design in two-dimensional sampling. Spatial separation of elements is considered;

thus this section of this work is not applicable to most problems of the type treated herein, since there is usually only one measurable dimension in the activity sampling problem. It is possible however to describe situations in which time and distance dimensions are present. The extension of Cochran's work is twofold. First, the conditions stated in Equations (44) are relaxed to the extent that each X_i of the finite population is considered as a random variable with mean μ_i and variance σ_i^2 ; μ_i is assumed to be distributed about μ with variance σ^2 , and $E(\mu_i - \mu)(\mu_j - \mu) = \rho_{ij}\sigma^2$, where ρ_u of Equations (44) is defined as follows:

$$\rho_u = \frac{1}{nk - u} \sum_{i=1}^{nk-u} \rho_{i \ i+u} . \quad (55)$$

In this case, the Equations (45)-(47) require the addition of the superposed variation,

$$\frac{1}{n} \left(1 - \frac{1}{k}\right) \frac{1}{nk} \sum_{i=1}^{nk} \sigma_i^2 , \quad (56)$$

to the right hand side. The second extension of Cochran's work consists of considering the process to be a continuous one. Equations (45)-(47) pass to the following forms:

$$\sigma_r^2 \sim \frac{\sigma^2}{n} , \quad (57)$$

$$\sigma_{st}^2 \sim \frac{\sigma^2}{n} \left[1 - \frac{2}{d^2} \int_0^d (d - u) \rho_u \, du \right] , \quad (58)$$

and

$$\sigma_{sy}^2 \sim \frac{\sigma^2}{n} \left[1 - \frac{2}{d} \int_0^{\infty} p_u \delta u + 2 \sum_{u=1}^{\infty} p_{du} \right], \quad (59)$$

where d is the interval between observations, and the first observation is randomly selected on $[0, d]$.

In 1950, Finney (56) extended his earlier report dealing with timber and forest surveys. Data were selected in which there was a marked period; and by comparing random, stratified random, and systematic sampling, the following conclusion, which seems justified, was given.

In general, systematic sampling was more precise than any form of random sampling at the same intensity, but this property was violently upset if the interval between successive strips of the systematic sample was a multiple of the period (56, p. 110).

Cochran included a chapter on the subject of systematic sampling in his excellent text (39) which has become one of the standard references on sampling. He first considers a particular finite population of $N = nk$ elements; and for a fixed value of k , x_{ij} denotes the value of the variable associated with the j^{th} element of the i^{th} possible systematic sample, and \bar{x}_i is the mean of the i^{th} possible systematic sample. The population mean is denoted by μ' , and inference is drawn with respect to the particular realization. The sampling design specifies the random selection of one of the possible k systematic samples, each having probability $1/k$ of being selected. The sample mean obtained in this manner, is shown to be an unbiased estimate of μ' , and it is denoted by \bar{x} (39, p. 162). The variance of the sample mean is shown to be $\frac{\sigma^2}{\bar{x}}$, where

$$\sigma_{\bar{x}}^2 = \sigma^2 - \frac{1}{nk} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2, \quad (60)$$

and

$$\sigma^2 = \frac{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \mu')^2}{N}. \quad (61)$$

An important corollary gives the necessary and sufficient conditions for the superiority of systematic sampling for the estimation of μ' for a given k and a particular finite population.

The mean of a systematic sample is more precise than the mean of a random sample if and only if

$$\frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 > \sigma^2 \quad (39, \text{p. 163}). \quad (62)$$

Cochran points out that this result, applicable to cluster sampling in general, states that systematic sampling is more precise than simple random sampling if "the variance within systematic samples is larger than the population variance as a whole" (39, p. 164). It is further noted that:

Systematic sampling is precise where units within the same sample are heterogeneous, and is imprecise when they are homogeneous. The result is obvious intuitively. If there is little variation within a systematic sample relative to that in the population, the successive units in the sample are repeating more or less the same information (39, p. 164).

Cochran notes that the previous literature of systematic sampling divides itself into two groups as that in which some function for the values x_i has been assumed and that in which a natural population has

been investigated. Four artificial populations types are considered as (1) populations in random order, (2) populations with linear trend, (3) populations with periodic variation, and (4) autocorrelated populations. Cases three and four have already been considered in this review, and Cochran shows that from a long-run viewpoint, the variances from systematic and random sampling have the same expected values in case one. He shows that systematic sampling is less precise than stratified random sampling but more precise than simple random sampling in case two.

In the case of stratified systematic sampling, an important result is obtained. The mean of the systematic sample in stratum h is denoted as $\bar{x}_{sy h}$, and it is an unbiased estimate of μ' , the population mean associated with the one realization or single population. The sample mean associated with this stratified systematic design is obtained by selecting L independent systematic samples, one from each stratum. The sample mean is denoted by $\bar{x}_{st sy}$, where

$$\bar{x}_{st sy} = \sum_{h=1}^L \frac{N_h}{N} \bar{x}_{sy h}, \quad (63)$$

and N_h/N is the fraction of elements in the h^{th} stratum. The variance of $\bar{x}_{st sy}$ is shown to be $\sigma_{\bar{x}_{st sy}}^2$, where

$$\sigma_{\bar{x}_{st sy}}^2 = \sum_{h=1}^L \left(\frac{N_h}{N} \right)^2 \sigma_{\bar{x}_{sy h}}^2. \quad (64)$$

Cochran develops an upper bound on the variance stated in Equation (64). He suggests however that the number of strata exceed 20 before it is employed.

$$\sigma_{\bar{x}_{st sy}}^2 = \sum_{\substack{\text{pairs}(h,j)\text{with} \\ \text{distinct} \\ \text{membership}}} \left(\frac{N_h}{N}\right)^2 (\bar{x}_{sy h} - \bar{x}_{sy j})^2 \quad (65)$$

The domain of summation consists of $L/2$ pairs formed from the L strata before sampling is accomplished. No two pairs have a common stratum. It is noted that this estimate

is on the average an overestimate, even if periodic effects are present within the strata. The amount of overestimation depends on terms in $(\mu'_h - \mu')$. So far as can be predicted, strata in the same pair should have about the same population means (39, p. 183).

Recent Developments in Systematic Sampling

In 1952, Jowett (82) reported an application of systematic sampling to the selection of coal specimens from a conveyor belt for ash content analysis. This paper does not seem to add significantly to the literature of systematic sampling.

In his most recent paper on systematic sampling, Madow (95) proved two theorems comparing the merits of centered systematic sampling with the first observation at the midpoint of $[0,d]$ and random start systematic sampling. The results indicate that the variability associated with the centered design is less when the correlogram is monotone decreasing and convex; however the bias associated with the centered design does not enhance the technique in so far as practical applicability

is concerned when little is known of the population behavior.

Williams (160) extended the results of Cochran's paper (38) in much the same manner as Quenoville (129). The form for the variance of systematic sampling (160, p. 138) is somewhat different from that obtained by Quenoville; however the result is essentially the same.

Jones (81) compared systematic sampling with multiple random starts to a design calling for the random selection of g elements from each group of k elements in the population, where there are $N = nk$ elements in the population. The latter design differs from stratified random sampling in the usual sense in that the first elements selected from each group form the first sample, the second elements form the second sample, and this combining of sample elements is continued until the g^{th} elements selected from each group form the g^{th} sample. In both cases the sample mean is calculated as indicated in Equation (37), and the estimate is shown to be an unbiased estimate of the mean μ' of the finite population. In both designs, an unbiased estimate of the variance of the sample mean was shown to be $\hat{\sigma}_{\bar{x}}^2$, where

$$\hat{\sigma}_{\bar{x}}^2 = \frac{(N - n)}{N} \frac{1}{g(g - 1)} \sum_{r=1}^g (\bar{x}_r - \bar{\bar{x}})^2 . \quad (66)$$

Gautschi (60) compared the sampling variances associated with systematic sampling with g random starts and n observations for each start to systematic sampling with a single random start and ng observations. For a monotone decreasing convex correlogram, the single random

start is shown to give an estimate having less variance. A significant concluding remark is given as follows:

When the statistician has the choice between systematic sampling and systematic sampling with multiple random starts, he is likely to use the latter procedure because its variance can be estimated from the sample and the estimate is unbiased whatever the form of the population (60, p. 394).

Hajek (65) extended the results of Cochran's theorem by relaxing the assumption requiring stationarity of the process mean and variance and replacing this assumption by the assumption that the coefficient of variation is stationary. The conclusions are the same for populations or processes having a monotone decreasing, convex correlogram.

Concluding Remarks

The measurement of uncertainty in systematic samples has been treated in two ways. Realistic population models have been constructed to account for the role of variability, and replication employing multiple random starts has been utilized. Both of these methods are employed in this thesis.

P A R T I I

CHAPTER IV

POPULATION DESCRIPTION AND SYSTEMATIC SAMPLING DESIGNS FOR SIMPLE ACTIVITY STRUCTURE

Introduction

A simple activity structure is defined as the activity of one subject, animate or inanimate, where all activity is classified as belonging to A , the state of interest, or A' (not A). A graphical representation corresponding to one realization of such a process over a period $[0, T]$ is shown in Figure 2. It is the objective of this chapter to describe quantitative models which replace the graphical model and to develop quantitative measures to describe the characteristics of processes as well as process realizations of simple activity. A secondary objective is the description of sampling designs for both single random starts and multiple random starts.

Population Description

Periodicity in the Population

If the lengths of the spans over which the activity is in states A' and A are constant for all occurrences of these states, then the activity will be called strictly periodic. It is convenient to denote the length of the span of the j^{th} occurrence of A by a_j and the length of the span of the j^{th} occurrence of A' by a'_j . In the case in which a'_j and a_j are mutually independent random variables for all

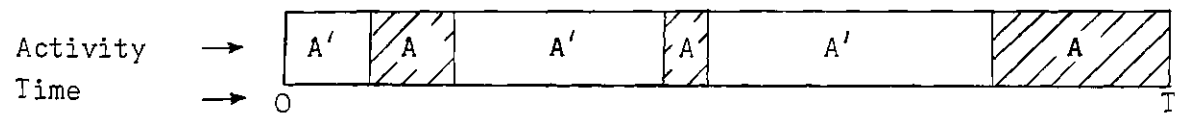


Figure 2. Simple Activity Structure

j , the activity is not strictly periodic. In this case, in which the densities of a'_j and a_j are negative exponential, the process or activity is called random, and the process behavior is Markovian. An infinite number of activity types exist between the random and strictly periodic types. The coefficient of variation of the cycle length is a measure of the degree of periodicity. The cycle length for the j^{th} cycle is denoted by $Y_j = a'_j + a_j$, and the coefficient of variation of this length is assumed independent of j . The coefficient of variation is denoted by $CV(Y_j)$, and it is the ratio of the standard deviation of Y_j to the mean of Y_j . If $CV(.)$ is small, the activity is called strongly periodic, and if its value is large, the activity is called weakly periodic. The symbol CV will be hereafter used for $CV(Y_j)$.

Infinite Population

We consider an infinite population of instants, t , on the interval $[0, T]$, and the graphical model of Figure 2 is replaced with the following quantitative model.

$x(t) = 1$; if the activity is in state A at time t .

$x(t) = 0$; if the activity is in state A' at time t .

The symbol $x(t)$ denotes a particular realization of $X(t)$, the corresponding stochastic process (or deterministic process in the case in which a'_j and a_j are constants).

If we consider the population to consist of only one realization of the process, the population mean and variance are respectively defined as

$$p_T = \frac{1}{T} \int_0^T x(t) dt, \quad (67)$$

and

$$\sigma_T^2 = \frac{1}{T} \int_0^T [x(t)]^2 dt - p_T^2. \quad (68)$$

This population variance may be interpreted as the variance of the mean of a random sample of size one selected from the instants on $[0, T]$. The serial correlation between observations separated by time u is given by $\rho_{T,u}$, where

$$\rho_{T,u} = \frac{1}{p_T(1 - p_T)} \left\{ \frac{1}{(T-u)} \int_0^{(T-u)} [x(t) - p_T][x(t+u) - p_T] dt \right\}. \quad (69)$$

In the case of process characteristics rather than the characteristics associated with one realization, the process is assumed to be a stationary stochastic process with mean and variance defined below:

$$p = E[X(t)] \quad (70)$$

and

$$\sigma^2 = E\{X(t) - p\}^2 = p(1 - p).$$

The correlation function for points t_i, t_j is given by $\rho(t_i, t_j)$, where

$$\rho(t_i, t_j) = \frac{1}{p(1 - p)} [Q(t_i, t_j) - p^2], \quad (71)$$

and

$$Q(t_i, t_j) = P\{X(t_i) = 1 \text{ and } X(t_j) = 1\}.$$

If the process is covariance stationary so that the correlation between $X(t)$ and $X(t + u)$ is a function of u only, then the correlation function is given by ρ_u , where

$$\rho_u = \frac{1}{p(1 - p)} [R(u) - p^2] \quad (72)$$

and

$$R(u) = E[X(t) X(t + u)] .$$

Finite Population

Although time is a continuous variable, there is only a finite number of intervals, $[0, v]$, $[v, 2v]$, ..., $[T - v, T]$, each of width v , in which observation can actually be made due to the limitations of human and mechanical observation devices. The realization $x(t)$ is not always constant across each interval, and changes of state will usually occur at some point within the interval; however the observation mechanism is such that $x(t)$ is observed as being constant on the small interval. This is true for any continuous random variable. In continuous stopwatch production study, changes are usually recorded to the nearest hundredth of a minute. This argument, coupled with the fact that the literature of survey sampling has dealt with finite populations, where several valuable theorems are available, leads to the consideration of a finite population of instants. From an engineering standpoint, this consideration is not obnoxious, and factors for finite population correction may generally be ignored since $(N - n) / (N - 1) \rightarrow 1$ as $N \rightarrow \infty$.

We consider a population with $N = nk$ instants, where N , n , and k are integers. Associated with the j^{th} instant there is a

variable X_j , so that we have X_1, X_2, \dots, X_N ; and a particular realization of the variables is x_1, x_2, \dots, x_N , where

$x_j = 1$; if the activity is in state A at the j^{th} instant.

$x_j = 0$; if the activity is in state A' at the j^{th} instant.

If we consider a particular finite population or one realization, the mean, variance, and serial correlation are as shown respectively below.

$$p_N = \frac{1}{N} \sum_{j=1}^N x_j , \quad (73)$$

$$\sigma_N^2 = \frac{1}{N} \sum_{j=1}^N (x_j - p_N)^2 = p_N(1 - p_N) , \quad (74)$$

and

$$\rho_{k\delta} = \frac{1}{\sigma_N^2} \frac{1}{k(n - \delta)} \sum_{j=1}^{k(n-\delta)} (x_j - p_N) (x_{j+k\delta} - p_N) , \quad (75)$$

where δ is an integer.

Following Cochran's approach (38), we now consider the variables X_j , $j = 1, 2, \dots, N$, and not just one realization. In this case, for the covariance stationary process, the process mean, variance, and serial correlation are respectively given below.

$$p = E[X_j] , \quad (76)$$

$$\sigma^2 = E[X_j - p]^2 = p(1 - p) , \quad (77)$$

and

$$\rho_u = \frac{1}{p(1 - p)} \{E[(X_i)(X_{i+u})] - p^2\} . \quad (78)$$

These relations are the correspondents of those given in Equations (70), (71), and (72).

Systematic Sampling Designs

One Random Start

First, we consider the systematic sampling of an infinite population. The period of study $[0, T]$ is divided into n intervals of width d , so that $nd = T$. An instant is randomly selected with a uniform distribution on $[0, d]$ and subsequent observations are spaced at intervals of width d apart following the first.

In the case of a finite population of size kn , one of the instants, $1, 2, \dots, k$, is randomly selected, and every k^{th} instant thereafter is selected. In both of these cases, the sample size is n .

Multiple Random Starts

In the case of the infinite population, g random observations are made on the interval $[0, d]$, and each of the instants selected serves as a starting point for a systematic sample selected in the manner described in the previous section. There are g systematic samples of size n each, and the total number of observations is ng .

In the finite population case, g of the instants, $1, 2, \dots, k$, are randomly selected, and each instant serves as a starting point for a systematic sample of n instants selected in the manner described in the previous section.

CHAPTER V

A SINGLE OCCURRENCE OF THE ACTIVITY

Introduction

In this chapter, we consider an activity, A , which occurs at most once during each time period such as a shift or half-shift. All other activity is called A' (not A). Sampling is employed to estimate " a ," the length of the span covered by A , or p_T , the fraction of the time period $[O,T]$ covered by activity A . Systematic sampling is to be shown to yield an estimate of " a " or p_T having less sampling variance than either simple random sampling or stratified random sampling. In 1948, Yates (164) outlined some mathematical methods for treating this type of process, and his methods are utilized in this chapter.

Examples of this activity type are numerous, and there is no restriction regarding the point of the time period at which the activity begins and ends; however there must be no more than one occurrence during the period. Some examples are:

1. The routine maintenance and adjustment of a machine which occurs once during each time period such as a shift or half-shift.
2. The utilization of an overhead crane, usually used for the movement of material in a production area, for moving material in a nonproduction area once per shift.

3. The daily removal of parts which have fallen into a plating tank.

4. The thorough washing of the mixer on concrete mixer trucks at the end of each day's use.

5. Clerical filing, where the clerk stops once each shift or half-shift to file.

6. The replenishment of raw stock, where the operator stops his usual work once during each period to replenish his stock.

7. The familiar coffee break, which occurs once per half-shift, whose length may nominally be 15 minutes, but in fact is a random variable.

Although several activities, not all conforming to the assumptions of this analysis, may be observed during the sampling procedure, the argument of this chapter is restricted to those activities following a pattern described in the preceding paragraphs.

The Mathematical Model of Sampling Designs

Consider a time period $[0, T]$ divided into n subperiods of length d , so that $nd = T$ as shown in Figure 3. The activity A begins at point t_b and ends at point t_e , where $0 \leq t_b \leq t_e \leq T$. A realization of the process $X(t)$ is denoted by $x(t)$, where

$$x(t) = 1 ; \quad t_b \leq t \leq t_e .$$

$$x(t) = 0 ; \quad \text{otherwise} .$$

Systematic sampling is to be considered first. The first

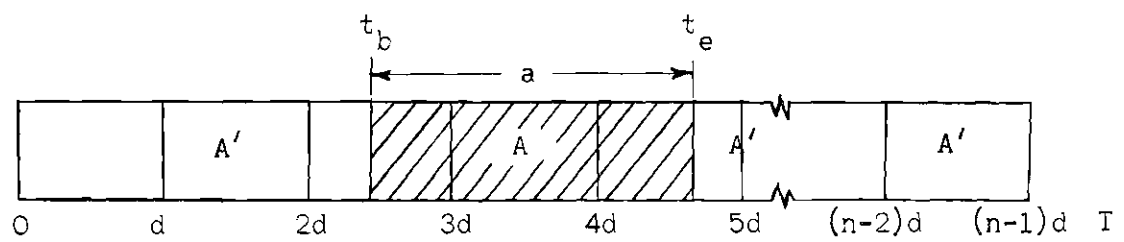


Figure 3. The Division of a Period into Intervals

observation is randomly located on the interval $[0, d]$ at point t_1 , and subsequent observations are spaced at times d units apart following t_1 .

Let

$$Z_{1 \text{ sy } d} = \sum_{j=0}^{n-1} x(t_1 + jd) , \quad (79)$$

and

$$\hat{p}_{1 \text{ sy } d} = \frac{d}{T} Z_{1 \text{ sy } d} = \frac{1}{n} Z_{1 \text{ sy } d} , \quad (80)$$

where $x(t_1 + jd) = 0, 1$ is the result of the $(j+1)^{\text{th}}$ observation made in the manner described above.

If $a < d$, $P(Z_{1 \text{ sy } d} = 1) = a/d$, and $P(Z_{1 \text{ sy } d} = 0) = 1 - a/d$; and it follows from the random start on $[0, d]$ that

$$E(Z_{1 \text{ sy } d}) = a/d ,$$

and

$$V(Z_{1 \text{ sy } d}) = (a/d) (1 - a/d) .$$

Likewise,

$$E(\hat{p}_{1 \text{ sy } d}) = \frac{1}{n} (a/d) ,$$

and

$$V(\hat{p}_{1 \text{ sy } d}) = \frac{1}{n^2} (a/d) (1 - a/d) .$$

Since the expected value and variance of $\hat{p}_{1 \text{ sy } d}$ are $1/n$ and $1/n^2$ multiples of the expected value and variance of $Z_{1 \text{ sy } d}$ respectively, the immediately following analysis will be restricted to the sums.

If $kd < a < (k+1)d$, where k is a positive integer,

$$E(Z_{1 \text{ sy } d}) = k + 1\left(\frac{a}{d} - k\right) = a/d ;$$

and

$$V(Z_{1 \text{ sy } d}) = c(1 - c) ,$$

where

$$c = (\frac{a}{d} - k)$$

is the fractional part of a/d . It is noted that there are at least k observations showing A and at most $(k + 1)$. Due to the random start on $[0, d]$, the activity will be observed $(k + 1)$ times with probability $(\frac{a}{d} - k)$. In summary for systematic sampling, we have

$$\begin{aligned} V(Z_{1 \text{ sy } d}) &\equiv V_{1 \text{ sy } d} = (a/d) (1 - a/d) ; \quad a < d , \quad (81) \\ &= c(1 - c) ; \quad kd < a < (k+1)d , \end{aligned}$$

where

$$c = (\frac{a}{d} - k) .$$

Stratified random sampling with one randomly located observation in each stratum, an interval of width d , is now considered. Let

$$Z_{1 \text{ str } d} = \sum_{j=1}^n x(t_j) , \quad (82)$$

where $x(t_j) = 0, 1$ is the result of the j^{th} observation taken at time t_j ; and t_j is randomly located on the interval of width d .

If $a < d$, there are two possibilities to consider. The first is where the span of activity, having length a , lies entirely within one interval. Since the end point of the span can fall anywhere on the interval, and it must be on the subinterval of width $(d - a)$ in

order for the span to fall wholly within the interval, $(1 - a/d)$ of the spans, $a < d$, will fall entirely within the interval in the long run if the intervals are randomly located with reference to the beginning point of the span. This assumption is incorporated in the analysis of the stratified designs. The second possibility here is for the span to fall in two intervals, and since either this or the first possibility must occur, $1 - (1 - a/d)$ of the spans will fall in two intervals in the long run.

In the first situation,

$$P(Z_{1 \text{ str } d} = 1) = a/d ,$$

and

$$P(Z_{1 \text{ str } d} = 0) = 1 - a/d ,$$

which gives

$$E(Z_{1 \text{ str } d}) = a/d ,$$

and

$$V(Z_{1 \text{ str } d}) = (a/d) (1 - a/d) .$$

Where a part of the span lies in each of two intervals, let m and $(a - m)$ represent the lengths of the activity span in the two intervals respectively, where m is assumed to have uniform density $1/a$ on the interval $[0, a]$. In this case, $Z_{1 \text{ str } d} = 0, 1, 2$; and

$$P(Z_{1 \text{ str } d} = 0) = (1 - m/d) (1 - a/d + m/d) ,$$

$$P(Z_{1 \text{ str } d} = 1) = (1 - m/d) (a/d - m/d) + (m/d) [(1 - a/d) + (m/d)]$$

and

$$P(Z_{1 \text{ str } d} = 2) = (m/d) (a/d - m/d) .$$

This gives:

$$\begin{aligned} E(Z_{1 \text{ str } d} | m) &= (1) \left[\left(1 - \frac{m}{d}\right) \left(\frac{a}{d} - \frac{m}{d}\right) + \left(\frac{m}{d}\right) \left(1 - \frac{a}{d} + \frac{m}{d}\right) \right] + (2) \left[\left(\frac{m}{d}\right) \left(\frac{a}{d} - \frac{m}{d}\right) \right] \\ &= a/d, \end{aligned}$$

and

$$V(Z_{1 \text{ str } d} | m) = \left[\frac{a}{d} - \frac{a^2}{d^2} + \frac{2am}{d^2} - \frac{2m^2}{d^2} \right].$$

Then

$$E(Z_{1 \text{ str } d}) = a/d,$$

and

$$\begin{aligned} V(Z_{1 \text{ str } d}) &= \int_0^a \left[\frac{a}{d} - \frac{a^2}{d^2} + \frac{2am}{d^2} - \frac{2m^2}{d^2} \right] \frac{1}{a} dm \\ &= a/d(1 - 2a/3d). \end{aligned}$$

Now applying the weights $(1 - a/d)$ and (a/d) to the values obtained for each case, we get, for $a < d$,

$$E(Z_{1 \text{ str } d}) = \left(\frac{a}{d}\right) \left(1 - \frac{a}{d}\right) + \left(\frac{a}{d}\right) \left(\frac{a}{d}\right) = \frac{a}{d},$$

and

$$\begin{aligned} V(Z_{1 \text{ str } d}) &= \left(\frac{a}{d}\right)^2 \left(1 - \frac{a}{d}\right)^2 + \frac{a^2}{d^2} \left(1 - \frac{2a}{3d}\right) \\ &= \left(\frac{a}{d}\right) \left(1 - \frac{a}{d}\right) + \frac{1}{3} \frac{a^3}{d^3}, \end{aligned}$$

where these operators are over all realizations of the process.

If $a > d$, there is only one situation to consider since the span cannot fall entirely within one interval. Let m and v represent the lengths of the span falling in the first and last partially filled intervals, respectively. If only two adjacent intervals contain the

entire span, and both are partially occupied, then $m + v = a$. If k intervals are fully occupied, then $dk + m + v = a$, where $k > 0$ is an integer. Combining these, we get $dk + m + v = a$; $k = 0, 1, 2, \dots$. The values m and v are assumed to have independent, uniform densities on $[0, d]$. In this case, $Z_1 \text{ str } d = k, k + 1, k + 2$; and

$$P(Z_1 \text{ str } d = k) = (1 - \frac{m}{d})(1 - \frac{v}{d}) ,$$

$$P(Z_1 \text{ str } d = k+1) = (\frac{m}{d})(1 - \frac{v}{d}) + (1 - \frac{m}{d})(\frac{v}{d}) ,$$

$$P(Z_1 \text{ str } d = k+2) = (\frac{m}{d})(\frac{v}{d}) ;$$

for $k = 0, 1, 2, \dots$. This gives

$$\begin{aligned} E(Z_1 \text{ str } d) &= \sum_{j=k}^{k+2} (j) P(Z_1 \text{ str } d = j) \\ &= k + \frac{m}{d} + \frac{v}{d} ; \end{aligned}$$

but since

$$k = \frac{a - m - v}{d} ,$$

$$E(Z_1 \text{ str } d) = a/d .$$

Since there is no variability contributed by the k fully occupied intervals,

$$V(Z_1 \text{ str } d) = (\frac{m}{d})(1 - \frac{m}{d}) + (\frac{v}{d})(1 - \frac{v}{d}) ,$$

and integrating with respect to m and v ,

$$\begin{aligned}
 V(Z_{1 \text{ str } d}) &= \int_0^d \int_0^d \left[\left(\frac{m}{d} \right) \left(1 - \frac{m}{d} \right) + \left(\frac{v}{d} \right) \left(1 - \frac{v}{d} \right) \right] \cdot \frac{1}{d} \cdot \frac{1}{d} \delta m \delta v \\
 &= 1/3 .
 \end{aligned}$$

In summary, for stratified sampling with one randomly located observation per stratum of width d , where the assumptions regarding the densities of m and v are satisfied, we have:

$$\begin{aligned}
 V(Z_{1 \text{ str } d}) &\equiv V_{1 \text{ str } d} = \left[\left(\frac{a}{d} \right) \left(1 - \frac{a}{d} \right) + \left(\frac{1}{3} \right) \left(\frac{a^3}{d^3} \right) \right]; \quad a \leq d, \quad (83) \\
 &= 1/3; \quad a \geq d .
 \end{aligned}$$

Stratified random sampling with two randomly located observations in each stratum, an interval of width $2d$, is now considered.

Let

$$Z_{2 \text{ str } 2d} = \sum_{j=1}^{n/2} \sum_{i=1}^2 x(t_{j_i}) ,$$

where $x(t_{j_i}) = 0,1$ is the result of the i^{th} observation in the j^{th} stratum, an interval of width $2d$.

If $a < 2d$, there are again two possibilities to consider. Logic similar to that employed in the consideration of the stratified design with one randomly located observation per stratum may be applied to lead to the consideration of two possibilities. Assumptions regarding the densities of m and v are the same as before except that the densities are $1/(2d)$ on the interval $[0, 2d]$. It follows that the entire span will fall entirely within the interval a fraction $1 - a/(2d)$, of the

time; and it will fall in two intervals $(a/2d)$ fraction of the time.

In the first instance, $Z_{2 \text{ str } 2d} = 0, 1, 2$; and

$$P(Z_{2 \text{ str } 2d} = 0) = \left(1 - \frac{a}{2d}\right)^2,$$

$$P(Z_{2 \text{ str } 2d} = 1) = 2\left(\frac{a}{2d}\right)\left(1 - \frac{a}{2d}\right),$$

$$P(Z_{2 \text{ str } 2d} = 2) = \left(\frac{a}{2d}\right)^2.$$

This gives

$$E(Z_{2 \text{ str } 2d}) = \sum_{j=0}^2 (j) P(Z_{2 \text{ str } 2d} = j) = a/d,$$

and

$$\begin{aligned} V(Z_{2 \text{ str } 2d}) &= \left[\sum_{j=0}^2 (j)^2 P(Z_{2 \text{ str } 2d} = j) \right] - \left[\frac{a}{d} \right]^2 \\ &= (2) \left(\frac{a}{2d}\right) \left(1 - \frac{a}{2d}\right). \end{aligned}$$

Where a part of the span, $a < 2d$, lies in each of two intervals, we again let m and $(a - m)$ represent the lengths of the span in the two intervals respectively; and $Z_{2 \text{ str } 2d} = 1, 2, 3, 4$ with probabilities:

$$\begin{aligned} P(Z_{2 \text{ str } 2d} = 0) &= \left(1 - \frac{m}{2d}\right)^2 \left[1 - \frac{(a - m)}{2d}\right]^2 \\ P(Z_{2 \text{ str } 2d} = 1) &= \left(1 - \frac{m}{2d}\right) \left[1 - \frac{(a - m)}{2d}\right]^2 \left(\frac{m}{2d}\right) \\ &\quad + \left(1 - \frac{m}{2d}\right)^2 \left[\frac{(a - m)}{2d}\right] \left[1 - \frac{(a - m)}{2d}\right] \end{aligned}$$

$$\begin{aligned}
P(Z_{2 \text{ str } 2d} = 2) &= \left(\frac{m}{2d}\right)^2 \left[1 - \frac{(a-m)}{2d}\right]^2 + \left(\frac{a-m}{2d}\right) \left[1 - \frac{m}{2d}\right]^2 \\
&+ \left(\frac{m}{2d}\right) \left(1 - \frac{m}{2d}\right) \left(\frac{a-m}{2d}\right) \left[1 - \frac{(a-m)}{2d}\right]
\end{aligned}$$

$$\begin{aligned}
P(Z_{2 \text{ str } 2d} = 3) &= \left(\frac{m}{2d}\right)^2 \left(\frac{a-m}{2d}\right) \left[1 - \frac{(a-m)}{2d}\right] \\
&+ \left(\frac{a-m}{2d}\right)^2 \left(\frac{m}{2d}\right) \left(1 - \frac{m}{2d}\right)
\end{aligned}$$

$$P(Z_{2 \text{ str } 2d} = 4) = \left(\frac{m}{2d}\right)^2 \left(\frac{a-m}{2d}\right)^2 .$$

We calculate

$$E(Z_{2 \text{ str } 2d} | m) = \sum_{j=0}^4 (j) P(Z_{2 \text{ str } 2d} = j) = a/d ,$$

and

$$\begin{aligned}
V(Z_{2 \text{ str } 2d} | m) &= \left[\sum_{j=0}^4 (j)^2 P(Z_{2 \text{ str } 2d} = j) \right] - \left[\frac{a}{d} \right]^2 \\
&= \left(\frac{m}{d}\right) \left(1 - \frac{m}{d}\right) + \left(\frac{a}{d} - \frac{m}{d}\right) \left(1 - \frac{a}{2d} + \frac{m}{2d}\right) .
\end{aligned}$$

Then averaging over possible values of m ,

$$E(Z_{2 \text{ str } 2d}) = a/d ,$$

and

$$\begin{aligned}
V(Z_{2 \text{ str } 2d}) &= \int_0^a \left[\left(\frac{m}{d}\right) \left(1 - \frac{m}{d}\right) + \left(\frac{a}{d} - \frac{m}{d}\right) \left(1 - \frac{a}{2d} + \frac{m}{2d}\right) \right] \cdot \frac{1}{a} dm \\
&= \left(\frac{a}{d} - \frac{a^2}{3d^2}\right) .
\end{aligned}$$

Applying the weights $(1 - \frac{a}{2d})$ and $(\frac{a}{2d})$ to the values obtained, we get for $a < 2d$,

$$E(Z_{2 \text{ str } 2d}) = a/d$$

and

$$V(Z_{2 \text{ str } 2d}) = \frac{1}{12} \frac{a}{d} [12 - \frac{6a}{d} + \frac{a^2}{d^2}] .$$

If $a > 2d$, the span cannot fall entirely within one interval, and we again let m and v represent the length of span falling in the first and last partially filled intervals respectively; so that $k(2d) + m + v = a$, where k intervals are completely filled. The sum, $Z_{2 \text{ str } 2d}$, takes values $2k, 2k + 1, 2k + 2, 2k + 3, 2k + 4$, with the following probabilities:

$$P(Z_{2 \text{ str } 2d} = 2k) = (1 - \frac{m}{2d})^2 (1 - \frac{v}{2d})^2$$

$$\begin{aligned} P(Z_{2 \text{ str } 2d} = 2k+1) &= (1 - \frac{m}{2d})(\frac{m}{2d})(1 - \frac{v}{2d}) \\ &+ (1 - \frac{m}{2d})^2(\frac{v}{2d})(1 - \frac{v}{2d}) \end{aligned}$$

$$\begin{aligned} P(Z_{2 \text{ str } 2d} = 2k+2) &= (\frac{m}{2d})^2(1 - \frac{v}{2d})^2 + (\frac{v}{2d})^2(1 - \frac{m}{2d})^2 \\ &+ (\frac{m}{2d})(1 - \frac{m}{2d})(\frac{v}{2d})(1 - \frac{v}{2d}) \end{aligned}$$

$$\begin{aligned} P(Z_{2 \text{ str } 2d} = 2k+3) &= (\frac{m}{2d})^2(\frac{v}{2d})(1 - \frac{v}{2d}) \\ &+ (\frac{m}{2d})(1 - \frac{m}{2d})(\frac{v}{2d})^2 \end{aligned}$$

$$P(Z_{2 \text{ str } 2d} = 2k+4) = (\frac{m}{2d})^2 (\frac{v}{2d})^2 .$$

This gives

$$E(Z_{2 \text{ str } 2d}) = k + \frac{m}{2d} + \frac{v}{2d},$$

but since

$$k = \frac{a}{2d} - \frac{m}{2d} - \frac{v}{2d},$$

$$E(Z_{2 \text{ str } 2d}) = a/d.$$

Again, there is no variability contributed by the k fully occupied intervals, so

$$\begin{aligned} V(Z_{2 \text{ str } 2d}) &= \int_0^{2d} \int_0^{2d} \left[(2) \left(\frac{m}{2d} \right) \left(1 - \frac{m}{2d} \right) \right. \\ &\quad \left. + (2) \left(\frac{v}{2d} \right) \left(1 - \frac{v}{2d} \right) \right] \frac{1}{2d} \cdot \frac{1}{2d} \delta v \delta m = 2/3. \end{aligned}$$

In summary, for stratified random sampling with two randomly located observations per stratum of width $2d$, we have:

$$\begin{aligned} V(Z_{2 \text{ str } 2d}) &\equiv V_{2 \text{ str } 2d} = \left(\frac{1}{12} \right) \left(\frac{a}{d} \right) \left[12 - \frac{6a}{d} + \frac{a^2}{d^2} \right]; \quad a \leq 2d \quad (85) \\ &= 2/3; \quad a \geq 2d. \end{aligned}$$

If this argument is extended by allowing id to approach T , then we approach simple random sampling on the period $[0, T]$, and when $i = n$, we have simple random sampling on the period. In this case, as in the systematic sampling case, no assumptions regarding the distributions of m and v are required, and the sampling may be considered to be on one realization of the process. In the case of simple random sampling, we let

$$Z_{n \text{ sr } T} = \sum_{j=1}^n x(t_j), \quad (86)$$

where $x(t_j) = 0, 1$ depending on whether or not the j^{th} observation at point t_j , randomly selected on $[0, T]$, shows A' or A respectively. Then,

$$E(Z_{n \text{ sr } T}) = \frac{na}{T} = a/d,$$

and

$$\begin{aligned} V(Z_{n \text{ sr } T}) &\equiv V_{\text{sr}} = n\left(\frac{a}{T}\right)\left(1 - \frac{a}{T}\right) \\ &= (a/d)\left(1 - \frac{a}{nd}\right). \end{aligned} \quad (87)$$

If $n = 1$, $V_{\text{sr}} = (a/d)(1 - a/d)$, and as $n \rightarrow \infty$, $V_{\text{sr}} \rightarrow (a/d)$. The variances for the four methods of sample selection are compared graphically in Figure 4. If $n < 4$, the horizontal scale would need restriction. It is noted that if $a/d < 0.1$, all methods are of approximately the same precision; but as a/d increases, systematic sampling becomes much more precise than the other forms of random sampling.

Comparison of Simple Random and Systematic Designs

The remainder of this chapter will be restricted to a comparison of systematic and simple random sampling, where the estimation of the fraction of time covered by activity A is the object of the sampling. One realization of the process constitutes the population; and p_T , the parameter to be estimated, is as defined in Equation (67). The estimates to be compared are

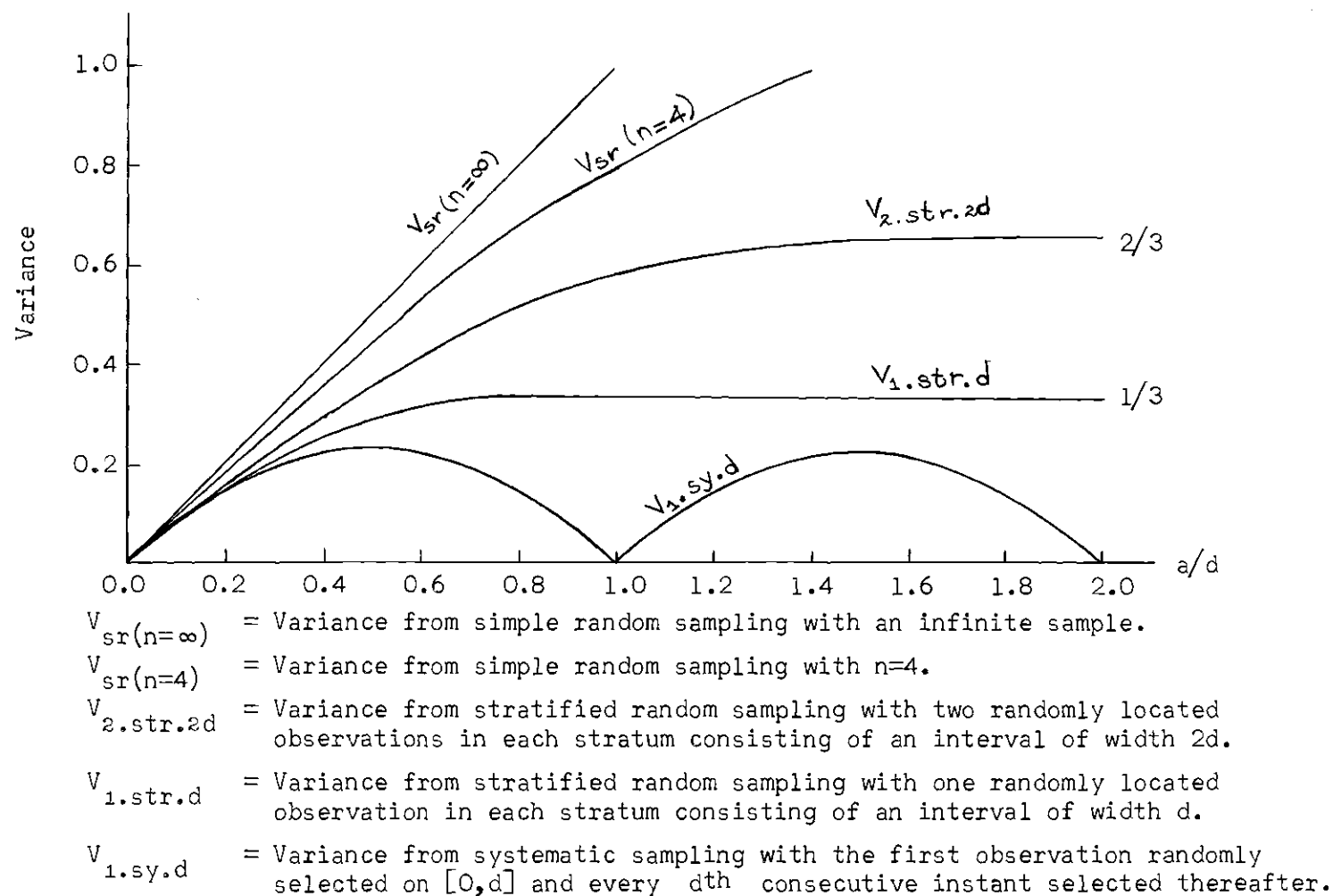


Figure 4. Sampling Variances for Different Methods of Sampling

$$\hat{p}_{1 \text{ sy } d} = \frac{1}{n}(Z_{1 \text{ sy } d}) ,$$

and

$$\hat{p}_{n \text{ sr } T} = \frac{1}{n} (Z_{n \text{ sr } T}) ,$$

for systematic and simple random designs respectively. The estimates are unbiased, since

$$E(\hat{p}_{1 \text{ sy } d}) = \frac{1}{n} (a/d) = a/T ,$$

and

$$E(\hat{p}_{n \text{ sr } T}) = \frac{1}{n} (a/d) = a/T .$$

The variances of the estimates are given below:

$$\begin{aligned} 1. \quad V(\hat{p}_{1 \text{ sy } d}) &= \frac{1}{n^2} [(a/d)(1 - a/d)] ; a < d \quad (88) \\ &= \frac{1}{n^2} [(c)(1 - c)] ; a \geq d , \end{aligned}$$

where $c = a/d - k$ is the fractional part of a/d .

$$2. \quad V(\hat{p}_{n \text{ sr } T}) = \frac{1}{n^2} \left(\frac{a}{d}\right) \left[1 - \left(\frac{1}{n}\right)\left(\frac{a}{d}\right)\right] . \quad (89)$$

It is noted that the variance of the estimate is uniformly less for systematic sampling when $n > 1$. An upper bound to the variance of $\hat{p}_{1 \text{ sy } d}$, which occurs at $a/d = 1/2$, is given by $V_{UB}(\hat{p}_{1 \text{ sy } d})$, where

$$V_{UB}(\hat{p}_{1 \text{ sy } d}) = (1/4)(1/n^2) . \quad (90)$$

The Distribution of $\hat{p}_{1 \text{ sy } d}$

The distribution of $\hat{p}_{1 \text{ sy } d}$ is graphically presented in Figure 5.

It is noted that the estimate takes one of two values in every case,

$$\text{Mean} = a/nd$$

$$\text{Variance} = (1/n^2)(C)(1 - C)$$

$$C = (a/d - k)$$

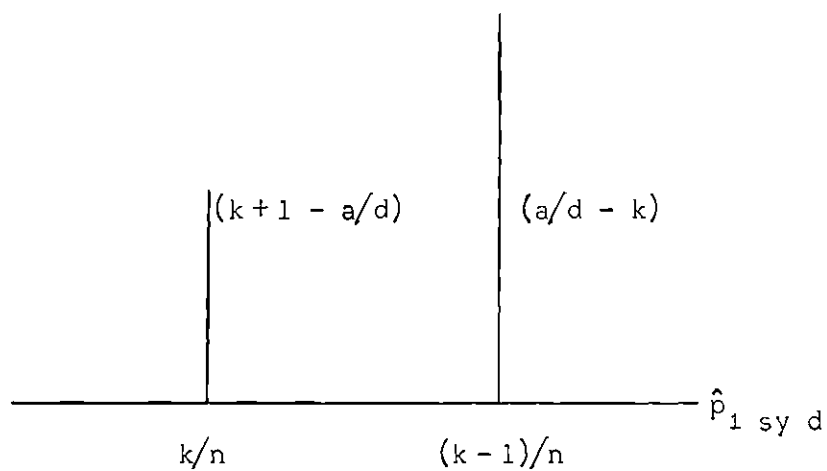


Figure 5. The Distribution of $\hat{p}_1 \text{ sy } d$ for
 $kd < a < (k+1)d$; $k = 0, 1, \dots, (n-1)$

where replication is over the same process realization; and a powerful confidence statement is an immediate consequence, although, for small n , the interval is rather wide.

$$P \left[\frac{(Z_1 \text{ sy } d) - 1}{n} \leq p_T \leq \frac{(Z_1 \text{ sy } d) + 1}{n} \right] = 1.0 \quad (91)$$

where the lower limit is truncated at zero when $Z_1 \text{ sy } d = 0$.

A simulation study was employed to test the theory developed in the previous section of this chapter. A single period, $[0, 480.0]$, was considered, and the three subperiods over which $x(t) = 0, 1, 0$, respectively, were $[0, 50.0]$, $[50.0, 111.0]$, and $[111.0, 480.0]$. Twelve values of d were employed, and there were 300 estimates calculated for each value of d . Each estimate resulted from a random selection of a starting point on $[0, d]$, and the summarized results are given in Table 1. The electronic computer program for this simulation is given in Table 11 of Appendix B.

Observations Over Several Time Periods

In practice, observations usually extend over several shifts, or half-shifts, each of which forms a time period $[0, T]$ as employed in the previous stages of this development. If the subscript h is employed for days, and we consider a day as forming a stratum, we have activity A with duration $(a)_1, (a)_2, \dots, (a)_h, \dots, (a)_L$ on days $1, 2, \dots, L$. We again assume that the sampling is of a particular realization of the process, so that we consider within period sampling error. Let the number of observations and the interval between observations be denoted by

Table 1. Comparison of Simulated and Theoretical Results for Systematic Sampling of a Period on Which the Activity Occurs Once, $T = 480$, $a = 61$, and $p = 0.1271$, 300 Replications for each Value of n .

| n | d | a/d | Random Sampling Theoretical | | Systematic Sampling Theoretical | | Systematic Sampling Simulation Expt. | | Lower Value LV=k/n | Upper Value UV = $\frac{(k+1)}{n}$ | Simulation Freq. LV. | Simulation Freq. UV. |
|-----|------|--------|-----------------------------|----------|---------------------------------|----------|--------------------------------------|------------------|-----------------------|---------------------------------------|----------------------|----------------------|
| | | | Mean | Std.Dev. | Mean | Std.Dev. | Sample Mean* | Sample Std.Dev.* | | | | |
| 192 | 2.5 | 24.400 | 0.1271 | 0.0240 | 0.1271 | 0.0026 | 0.1274 | 0.0026 | 0.1250 | 0.1302 | 164 | 136 |
| 96 | 5.0 | 12.200 | 0.1271 | 0.0340 | 0.1271 | 0.0042 | 0.1275 | 0.0045 | 0.1250 | 0.1354 | 227 | 73 |
| 64 | 7.5 | 8.133 | 0.1271 | 0.0416 | 0.1271 | 0.0053 | 0.1275 | 0.0057 | 0.1250 | 0.1406 | 252 | 48 |
| 48 | 10.0 | 6.100 | 0.1271 | 0.0481 | 0.1271 | 0.0063 | 0.1269 | 0.0060 | 0.1250 | 0.1458 | 273 | 27 |
| 32 | 15.0 | 4.066 | 0.1271 | 0.0589 | 0.1271 | 0.0078 | 0.1275 | 0.0085 | 0.1250 | 0.1563 | 276 | 24 |
| 24 | 20.0 | 3.050 | 0.1271 | 0.0680 | 0.1271 | 0.0091 | 0.1269 | 0.0088 | 0.1250 | 0.1667 | 286 | 14 |
| 16 | 30.0 | 2.033 | 0.1271 | 0.0833 | 0.1271 | 0.0112 | 0.1267 | 0.0101 | 0.1250 | 0.1875 | 292 | 8 |
| 12 | 37.5 | 1.626 | 0.1271 | 0.0962 | 0.1271 | 0.0403 | 0.1331 | 0.0409 | 0.0833 | 0.1667 | 121 | 179 |
| 10 | 45.0 | 1.355 | 0.1271 | 0.1053 | 0.1271 | 0.0479 | 0.1347 | 0.0476 | 0.1000 | 0.2000 | 196 | 104 |
| 9 | 52.5 | 1.162 | 0.1271 | 0.1110 | 0.1271 | 0.0409 | 0.1285 | 0.0404 | 0.1111 | 0.2222 | 253 | 47 |
| 8 | 60.0 | 1.016 | 0.1271 | 0.1178 | 0.1271 | 0.0160 | 0.1267 | 0.0143 | 0.1250 | 0.2500 | 296 | 4 |
| 6 | 80.0 | 0.763 | 0.1271 | 0.1360 | 0.1271 | 0.0709 | 0.1267 | 0.0712 | 0.0000 | 0.1667 | 72 | 228 |

*From the empirical distribution corresponding to the theoretical distribution of Figure 5.

n and d respectively for days $1, 2, \dots, L$; and there is to be an independent random start on $[0, d]$ for each day. Let the total length of span of activity A for the L days be denoted by $a.$, where

$$a. = \sum_{h=1}^L (a)_h, \quad (92)$$

so that the parameter to be estimated is p_{LT} , where

$$p_{LT} = \frac{a.}{LT}. \quad (93)$$

It follows that an unbiased estimate of p_{LT} is \hat{p}_{LT} , where

$$\hat{p}_{LT} = \frac{1}{L} \sum_{h=1}^L \hat{p}_{1 \text{ sy } d}, \quad (94)$$

and an upper bound to the variance of \hat{p}_{LT} is $V_{UB}(\hat{p}_{LT})$, where

$$V_{UB}(\hat{p}_{LT}) = \frac{1}{4L^2} \sum_{h=1}^L \frac{1}{n^2} = \frac{1}{4Ln^2}. \quad (95)$$

If $(Z_{1 \text{ sy } d})_h$ is the sum defined in Equation (79) for day h , a confidence interval on p_{LT} is given as follows

$$P \left\{ \frac{\left[\sum_{h=1}^L (Z_{1 \text{ sy } d})_h \right] - L}{Ln} \leq p_{LT} \leq \frac{\left[\sum_{h=1}^L (Z_{1 \text{ sy } d})_h \right] + L}{Ln} \right\} = 1. \quad (96)$$

If a prediction interval on p_T , the fraction of time spent on activity A on some future day, is desired, and the random variable

$(\alpha)_h$, having realization a_h , has a stable distribution, variance components analysis may be employed to determine the day to day component of variance. The prediction interval is wider due to the necessity of accounting for the day to day component of variance.

Concluding Remarks

In concluding this chapter, it is noted that systematic sampling is uniformly more precise than simple random sampling when the activity occurs at most once per time period. It is further noted that an estimate of the variance of \hat{p}_T or \hat{p}_{LT} is not possible at this point from the sample; however the availability of an upper bound to the variance is helpful, and the nature of the distribution of the estimate provides for the making of rather powerful confidence statements on the parameter estimated. Estimation techniques for the variance of \hat{p}_T or \hat{p}_{LT} are presented in Chapter IX.

CHAPTER VI

MULTIPLE OCCURRENCE OF THE ACTIVITY ON THE PERIOD $[0,T]$ WITH SEVERE RESTRICTIONS ON THE DISTRIBUTIONS OF SPAN LENGTHS

Introduction

In this chapter, we consider an activity, A , which may occur more than once during the time period $[0,T]$; and here we let a_j represent the length of the j^{th} span of activity A , and a_j' is the symbol for the length of the j^{th} span of A' . Systematic sampling with a single random start on $[0,d]$ is to be employed to estimate "a" or $p_T = a/T$, where

$$a = \sum_{j=1}^{N'} a_j, \quad (97)$$

and there are N' occurrences of activity A .

The density of a_j is denoted by $f_1(a_j)$, and the density of a_j' is denoted by $f_2(a_j')$, where a_j and a_j' are considered as random variables. This chapter treats the case where these densities are truncated on the left, and the intervals $[\alpha, \infty)$ and $[\alpha', \infty)$ contain all of the measure for $f_1(.)$ and $f_2(.)$ respectively, where $\alpha > d$ and $\alpha' > d$. This case was first considered by Davidson, Hines, and Newberry in 1960 (47). The objective of this earlier work was to develop a formula for the variance of the estimate of p_T . The purpose was to show

that camera analysis with a fixed interval between frames provides a satisfactory estimate of p_T . This presentation extends the results obtained by Davidson et al. with a clarification and justification of the assumptions.

From an engineering standpoint, this case is applicable whenever there is some a priori knowledge indicating that $\alpha > d$ and $\alpha' > d$; and otherwise, if estimates, $\hat{\alpha}$ and $\hat{\alpha}'$, arrived at from study of historical data on similar activities, physical limitations of the activity (process), or judgement, are such that there is high confidence that $\alpha < \hat{\alpha}$ and $\alpha' < \hat{\alpha}'$.

This case is not applicable when a_j is the time required to perform a short element of a repetitive work cycle, since the observation density would be such that production study would be approached in order to satisfy the assumptions. If a number of long cycles are sequentially performed and followed by an idle span; and if the assumptions of the preceding paragraphs are satisfied, then this model would be applicable for the estimation of either the fraction of work time or the fraction of idle time, or the elemental times. The disadvantages associated with time delay photography incorporating one frame for each observation (128, p. 44) have been overcome somewhat by the memo-activity camera developed by H. G. Thuesen under National Science Foundation Grant No. 17674. This camera is capable of taking bursts of ten frames every two minutes, an interval which has been reported in several systematic sampling studies. The burst of frames provides for a more accurate classification of activity by the film analyst.

In practice, the assumptions underlying this case are such that a rather high sampling intensity must usually be utilized; however there are some types of activity where α and α' are large enough to permit a modest sampling intensity. Some examples of such activity are given below.

1. Any repetitive work cycle with "long" elements.
2. The operation of a process where equipment must be cleaned between the processing of batches.
3. The performance of surgery by a physician.
4. The utilization of commercial aircraft.

The Mathematical Model

Consider a time period $[0, T]$ divided into n subperiods of length d , so that $T = nd$. The j^{th} occurrence of activity A begins at time b_j and ends at time e_j , where $0 \leq b_j < e_j \leq T$ for $j = 1, 2, \dots, N'$. The j^{th} occurrence of A' begins at b'_j and ends at e'_j , where $b'_j < e'_j$. The process variable $X(t) = x(t)$, where

$$\begin{aligned} x(t) &= 1 ; & b_j \leq t \leq e_j ; & \text{ for } j = 1, 2, \dots, N' \\ &= 0 ; & b'_j \leq t \leq e'_j ; & \text{ for } j = 1, 2, \dots, N' , \end{aligned}$$

which describes the stochastic process since $a_j = e_j - b_j$ and $a'_j = e'_j - b'_j$ are random variables with densities $f_1(\cdot)$ and $f_2(\cdot)$ respectively.

Systematic sampling is considered with one random start on $[0, d]$ at point t_1 . Let

$$Z_{sy\ d} = \sum_{j=1}^n x[t_1 + (j-1)d] , \quad (98)$$

and

$$p_{sy\ d} = \frac{1}{n} [Z_{sy\ d}] . \quad (99)$$

Consider the sample vector $\{x(t_1), x(t_2), \dots, x(t_n)\}$ whose elements are 0's and 1's. Associated with each adjacent unequal pair of members of the sample vector, there is a change in activity and a value y_ℓ (for the ℓ^{th} such pair) so that:

$$a = d Z_{sy\ d} + \sum_{\ell=1}^K y_\ell , \quad (100)$$

where y_ℓ is a random variable on $[-d/2, d/2]$, and it is defined as follows: y_ℓ = the time span from the midpoint of an interval of width d , between two successive observations indicating a change in activity, to the actual point of change. The span y_ℓ is measured as positive toward the $x(t_i) = 0$, that precedes or follows the change. This situation is graphically illustrated in Figure 1 (See page 26).

Estimation and the Error of Estimate

The estimate of a to be considered is denoted by \hat{a} , where

$$\hat{a} = d Z_{sy\ d} . \quad (101)$$

The error of estimate is $(a - \hat{a})$ and this error is denoted by ξ ,

where

$$\xi = \sum_{\ell=1}^K y_\ell . \quad (102)$$

The value $K = 2N' - 1$ if $x(0) = 0$ and $x(T) = 1$, $K = 2N'$ if $x(0) = 0$ and $x(T) = 0$, and $K = 2N' - 2$ if $x(0) = 1$ and $x(T) = 1$.

The mean and variance of the error of estimate are given by

$$E(\xi) = \sum_{\ell=1}^K E(y_{\ell}), \quad (103)$$

and

$$V(\xi) = \sum_{\ell=1}^K V(y_{\ell}), \quad (104)$$

where the y_{ℓ} are assumed to be independent random variables.

Let

$$\phi_{\ell} = \sum_{j=1}^{[\ell/2]} (a'_j + a_j) + \delta a'_{[\ell/2]+1}$$

where $\delta = 0$ if ℓ is even and $\delta = 1$ if ℓ is odd. It is assumed that $K = 2N'$ and that a_j and a'_j are mutually independent random variables for $j = 1, 2, \dots, N'$. The density of ϕ_{ℓ} is denoted by $g_{\ell}(\cdot)$, and the density of y_{ℓ} is a conditional density that may be expressed as $h_{\ell}(\cdot)$, where

$$\begin{aligned} h_{\ell}(y_{\ell}) &= \frac{g_{\ell}(c_{\ell} + d/2 - y_{\ell})}{\int_{c_{\ell}}^{(c_{\ell}+d)} g_{\ell}(\phi_{\ell}) \delta \phi_{\ell}} ; \quad -d/2 \leq y_{\ell} \leq d/2, \quad \ell \text{ odd} \quad (106) \\ &= \frac{g_{\ell}(c_{\ell} + d/2 + y_{\ell})}{\int_{c_{\ell}}^{(c_{\ell}+d)} g_{\ell}(\phi_{\ell}) \delta \phi_{\ell}} ; \quad -d/2 \leq y_{\ell} \leq d/2, \quad \ell \text{ even;} \end{aligned}$$

and c_ℓ is the time at which the observation preceding the change in activity is made. The mean and variance of y_ℓ are given by $E(y_\ell)$ and $V(y_\ell)$, where

$$E(y_\ell) = \int_{-d/2}^{d/2} y_\ell h_\ell(s) ds \quad (107)$$

and

$$V(y_\ell) = \int_{-d/2}^{d/2} (y_\ell)^2 h_\ell(s) ds - [E(y_\ell)]^2 \quad (108)$$

An upper bound for the variance of ξ , defined in Equation (104), is desired. The approach to be used is not exact; however it is felt that good approximate results of the form to be given are of value. First we consider the distribution of φ_ℓ for large ℓ . Employing the central limit theorem,

$$g_\ell(\varphi_\ell) \sim N \{ \mu_{\varphi_\ell}, \sigma_{\varphi_\ell}^2 \} \text{ as } \ell \rightarrow \infty; \quad (109)$$

and

$$\mu_{\varphi_\ell} = \{ [\ell/2] (\mu_{a'} + \mu_a) + (\delta) \mu_{a'} \},$$

while

$$\sigma_{\varphi_\ell}^2 = \{ [\ell/2] (\sigma_{a'}^2 + \sigma_a^2) + (\delta) \sigma_{a'}^2 \},$$

and $\delta = 0$ for even ℓ , and $\delta = 1$ for odd ℓ . The values $\mu_{a'}$, $\sigma_{a'}^2$, μ_a , and σ_a^2 are the means and variances of the variables a'_j and a_j respectively. In this case the function $h_\ell(\cdot)$ has the greatest departure from a uniform distribution when $c_\ell = \mu_{\varphi_\ell} + \sigma_{\varphi_\ell} - d/2$ or $c_\ell = \mu_{\varphi_\ell} - \sigma_{\varphi_\ell} - d/2$, since the inflection points of the density $h_\ell(\cdot)$

are located at $\varphi_\ell = \mu_{\varphi_\ell} - \sigma_{\varphi_\ell}$, and $\varphi_\ell = \mu_{\varphi_\ell} + \sigma_{\varphi_\ell}$. Dividing the interval $[\mu_{\varphi_\ell} - \sigma_{\varphi_\ell} - d/2, \mu_{\varphi_\ell} + \sigma_{\varphi_\ell} + d/2]$ into range quartiles, the probability measure on the r^{th} range quartile is given by

$$P \{L_r \leq \varphi_\ell \leq U_r\} = \theta_r, \quad (110)$$

where L_r , U_r , and θ_r are the upper and lower limits to the r^{th} range quartile and the probability on $[L_r, U_r]$ respectively. If $(d)/(4\sigma_{\varphi_\ell}) = 0.05$, the probabilities are given in Table 2. The assumptions of Davidson, Hines, and Newberry were incorrect in regard to the distribution of y_ℓ ; however, for large ℓ , the distribution of y_ℓ is nearly uniform, and the fact that $E(y_\ell) \neq 0$ is not as serious as it seems since $E(y_\ell) > 0$ for odd ℓ , and $E(y_\ell) < 0$ for even ℓ , so that $E(\xi) \approx 0$. An upper bound to the variance is approximated by the following relation for large ℓ :

$$V(y_\ell) = d^2/12, \quad (111)$$

which gives

$$V(\xi) = (d^2 N')/6. \quad (112)$$

For small values of ℓ , the density $g_\ell(\cdot)$ may be quite non-normal depending on the densities of a_j and a'_j . We select densities which are adverse from this standpoint. Such densities would, for example, be described by exponential density functions truncated on the left so that the assumptions of this chapter are satisfied. Let

$$\begin{aligned} f_1(a_j) &= \lambda e^{-\lambda(a_j - \alpha)}; & a_j > \alpha \\ &= 0; & \text{otherwise,} \end{aligned} \quad (113)$$

Table 2. Probability on the Range Quartiles

| Interval | Probability | |
|------------------------------|-------------|-------------|
| | Odd ℓ | Even ℓ |
| $-d/2 \leq y_\ell \leq -d/4$ | 0.2686 | 0.2314 |
| $-d/4 \leq y_\ell \leq 0$ | 0.2562 | 0.2438 |
| $0 \leq y_\ell \leq d/4$ | 0.2438 | 0.2562 |
| $d/4 \leq y_\ell \leq d/2$ | 0.2314 | 0.2686 |

and

$$f_2(a_j') = \lambda e^{-\lambda'(a_j' - \alpha')} ; \quad a_j' > \alpha' \\ = 0 ; \quad \text{otherwise .}$$

Only the first two changes are to be considered with the corresponding density of y_1 and y_2 developed. The densities of φ_1 and φ_2 are:

$$g_1(\varphi_1) = \lambda' e^{-\lambda'(\varphi_1 - \alpha')} ; \quad \varphi_1 \geq \alpha' \quad (114) \\ = 0 ; \quad \text{otherwise ,}$$

and

$$g_2(\varphi_2) = \left\{ \left[\frac{\lambda' \lambda}{\lambda' - \lambda} \cdot e^{\lambda(\alpha + \alpha')} \right] e^{-\lambda \varphi_2} - \left[\frac{\lambda' \lambda}{\lambda' - \lambda} \cdot e^{\lambda'(\alpha' + \alpha)} \right] e^{-\lambda' \varphi_2} \right\} ; \\ \varphi_2 \geq \alpha' + \alpha \\ = 0 ; \quad \text{otherwise .}$$

The conditional densities for y_1 and y_2 are:

$$h_1(y_1) = \left[\frac{\lambda'}{e^{-\lambda'(d/2)} (1 - e^{-\lambda'd})} \right] e^{-\lambda' y_1} ; \quad -d/2 \leq y_1 \leq d/2 \quad (115)$$

$$h_1(y_1) = 0 ; \quad \text{otherwise ,}$$

and

$$h_2(y_2) = \left\{ \frac{\left[\lambda' \lambda e^{\lambda(\alpha' + \alpha - c_2 - d/2)} \right] e^{-\lambda y_2} - \left[\lambda' \lambda e^{\lambda'(\alpha' + \alpha - c_2 - d/2)} \right] e^{-\lambda' y_2}}{\left[\lambda' e^{\lambda(\alpha' + \alpha - c_2)} (1 - e^{-\lambda d}) \right] + \left[\lambda e^{\lambda'(\alpha' + \alpha - c_2)} (1 - e^{-\lambda' d}) \right]} \right\} ;$$

$$-d/2 \leq y_2 \leq d/2$$

$$h_2(y_2) = 0 ; \quad \text{otherwise .}$$

The mean and variance of y_1 are:

$$E(y_1) = \frac{1}{\lambda'} e^{\lambda' d} + \left(\frac{e^{\lambda' d} + 1}{e^{-\lambda' d} - 1} \right) \cdot \frac{d}{2}$$

and

$$\begin{aligned} V(y_1) &= \left(\frac{1}{\lambda'} \right)^2 \left[\frac{2(e^{\lambda' d} - 1)}{(1 - e^{-\lambda' d})} - e^{2\lambda' d} \right] \\ &+ \left(\frac{d}{\lambda'} \right) \left[\frac{(e^{\lambda' d} + 1)}{(e^{-\lambda' d} - 1)} \cdot (1 - e^{\lambda' d}) \right] \\ &+ \left(\frac{d^2}{4} \right) \left[\frac{(e^{\lambda' d} - 1)}{(1 - e^{-\lambda' d})} - \left(\frac{e^{\lambda' d} + 1}{1 - e^{-\lambda' d}} \right)^2 \right] \end{aligned} \quad (117)$$

Since $\lambda' d < 1$ because $1/\lambda' > \alpha > d$, $V(y_1) < d^2/12$. This result follows from algebraic simplification, and since $V(y_1) \leq V(y_2) \leq V(y_3) \dots \leq V(y_{2N'})$, an upper bound to $V(\xi)$ is given by

$$V(\xi) \leq [V(y_{2N'})] (2N') \quad (118)$$

$$\leq (d^2/12) (2N') .$$

We now consider the estimate of p_T denoted by $\hat{p}_{sy\ d}$. An upper bound to the variance of $\hat{p}_{sy\ d}$ is given by

$$V_{UB}(\hat{p}_{sy\ d}) = (1/n^2)(N'/6) , \quad (119)$$

since

$$\hat{p}_{sy\ d} = \hat{a}/(nd) .$$

A confidence statement on p_T may be constructed by employing Tchebycheff's inequality so that

$$\begin{aligned} P\{\hat{p}_{sy\ d} - (k/n) (\sqrt{N'/6}) \leq p_T \leq \hat{p}_{sy\ d} \\ + (k/n) (\sqrt{N'/6})\} \geq 1 - (1/k^2) \end{aligned} \quad (120)$$

In the case where observation is to extend over L strata such as days and a day forms a time period referred to in this development, $(a)_h$ denotes the sum of the values a_j for day h . The case is considered in which there are n observations per day, so that

$$p_{LT} = \frac{\sum_{h=1}^L (a)_h}{LT} \quad (121)$$

is the fraction of time covered by activity A during the L days.

An unbiased estimate and the upper bound to the variance of this estimate are given below:

$$\hat{p}_{LT} = \frac{1}{L} \sum_{h=1}^L (\hat{p}_{sy\ d})_h \quad (122)$$

$$V_{UB}(\hat{p}_{LT}) = \frac{1}{L^2} \left(\frac{1}{n^2}\right) \sum_{h=1}^L (N'_h/6) . \quad (123)$$

The symbols $(\hat{p}_{sy\ d})_h$ and N'_h represent the estimate for the h^{th} day and the number of occurrences of activity A on the h^{th} day.

A confidence statement on p_{LT} is given below:

$$P \{ \hat{p}_{LT} - k \sqrt{V_{UB}(\hat{p}_{LT})} \leq p_{LT} \leq \hat{p}_{LT} + k \sqrt{V_{UB}(\hat{p}_{LT})} \} \geq 1 - \frac{1}{k^2} . \quad (124)$$

A prediction interval for use on some future day should account for the day to day variability. If we consider the stochastic process $X(t)$ to be stationary, variance components analysis may be employed to estimate this component of variation.

Concluding Remarks

In conclusion, we compare simple random sampling and systematic sampling where the sample size n is fixed and constant for both designs. Considering one period $[0, T]$, systematic sampling is superior to random sampling if $N' < 6n(p)(1 - p)$. If there is no knowledge of p , we may consider the largest possible value of $(p)(1 - p)$, i.e., $1/2(1 - 1/2)$, in which case it follows that systematic sampling is superior if $N' < (3/2)(n)$.

CHAPTER VII

MULTIPLE OCCURRENCE OF THE ACTIVITY ON THE PERIOD $[0, T]$

Introduction

Many activities encountered by the practitioner do not qualify for treatment under the assumptions of the previous two chapters. This chapter treats the case where the activity of interest may occur more than once per time period, and there is no knowledge that the distribution of span lengths is truncated at some value greater than zero.

A covariance stationary process is assumed so that Equations (70), (71), and (72) describe the process characteristics for the infinite population concept, and Equations (76), (77), and (78) describe the process characteristics for the finite population concept.

Exponential Distributions of the Span Lengths of A and A'

There are numerous activities that behave in such a manner that the span length between occurrences of the activity is a random variable having a negative exponential probability distribution. Such durations are usually referred to as random, and some examples of activity that might be considered as satisfying the assumption are:

1. Machine activity, where the state is classified as either down for repair or operating.
2. The activity of telephone operators, where the state is classified as either idle or busy.

3. Irregular work of a servicing nature such as that described by Haines (64).

The variables a_j and a'_j are assumed to be mutually independent random variables having densities that are negative exponential, so that the process is Markovian (49, p. 117). The densities of a_j and a'_j are given below.

$$\begin{aligned} f_1(a_j) &= \lambda e^{-\lambda a_j}; & a_j \geq 0 \\ &= 0; & \text{otherwise} \\ f_2(a'_j) &= \lambda' e^{-\lambda' a'_j}; & a'_j \geq 0 \\ &= 0; & \text{otherwise} . \end{aligned} \tag{125}$$

If the process is covariance stationary as assumed, then

$$\rho_u = e^{-(\text{const.})|u|} \tag{126}$$

where

$$\text{const.} = \frac{1}{\left(\frac{\lambda' + \lambda}{\lambda \lambda'}\right) (p) (1 - p)} \tag{127}$$

A systematic sampling design with one random start yields an estimate of the mean having less variance than any form of random sampling at the same intensity. This statement is an immediate consequence from Cochran's Theorem, since the correlogram is monotone decreasing, convex (38, p. 117; 49, p. 117). The correlogram of Equation (126) is graphically shown as the solid line in Figure 6. for the case where $\lambda = \lambda' = 0.1$.

It is carefully noted that the variance of the estimate cannot be estimated from the sample; however, if the practitioner is satisfied

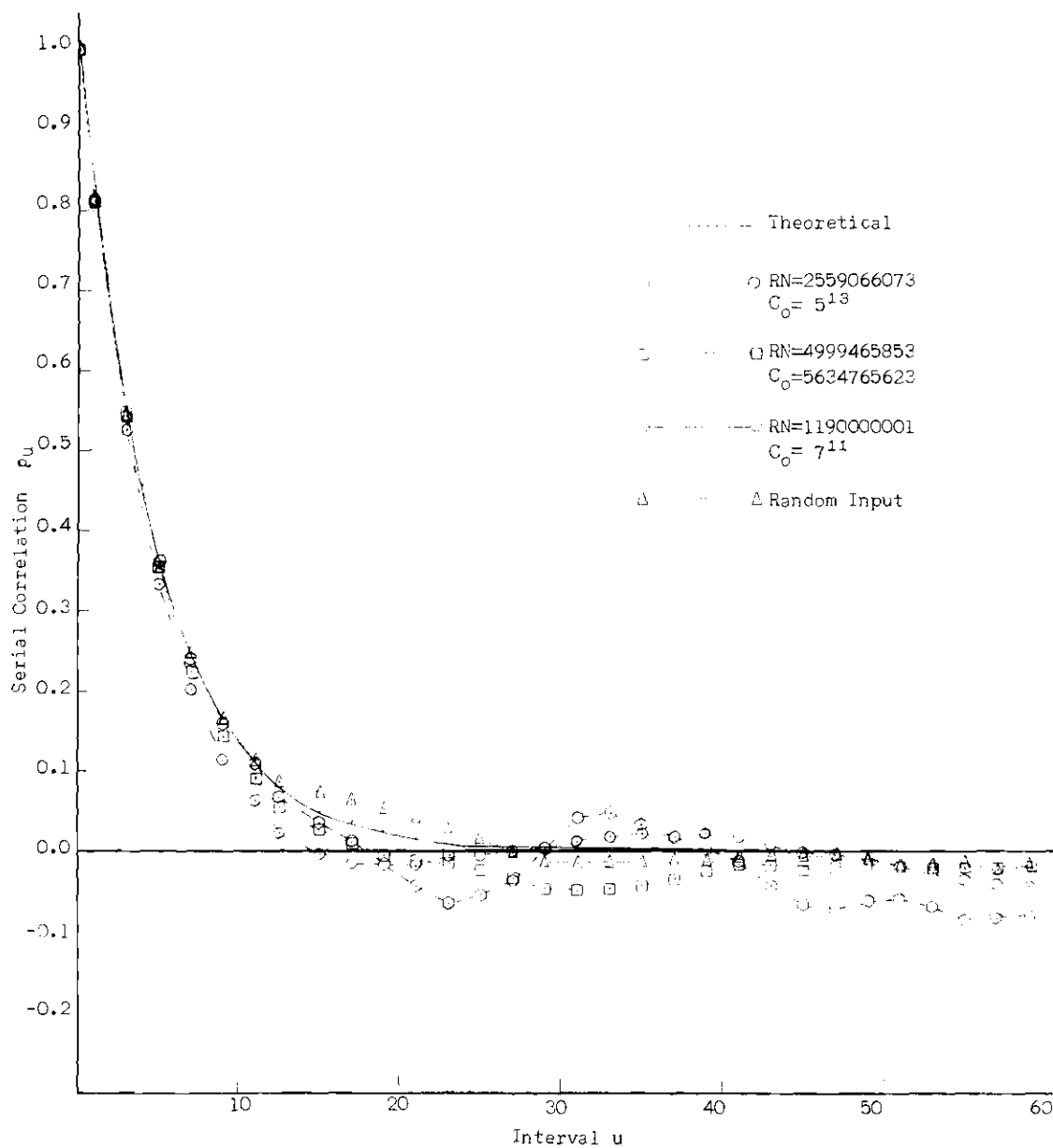


Figure 6. Correlation Functions from Simulated Process Realizations, and Theoretical Values, Where $\lambda = \lambda' = 0$, and $r = r' = 1$

to obtain the best result possible for a fixed sample size, and no confidence statement is required, then systematic sampling is optimal for this type of activity structure.

Activity sampling has been employed in the estimation of parameters to be utilized in solving queueing models in which the densities employed are, in many cases, negative exponential densities. Allderige (12) reported such an application in which random sampling was employed as the sampling design. In the case where exponential behavior is indicated, it is suggested that systematic sampling designs should be employed.

Other Distributions of the Span Lengths of A and A'

Introduction

There are numerous densities that might be considered for the random variables a_j and a_j' . In practice these variables may not have stable distributions, and in many cases, the variables may have given densities over subperiods of the entire period, and other densities over other subperiods; however in this section, we consider the case where the activity is periodic, and the densities are stable over the entire period. Two cases are considered, and in the first case, the densities of a_j and a_j' are assumed to be gamma densities. There are several reasons for this selection. First, this is a logical departure from the exponential in the direction of a decreased coefficient of variation and more central tendency. Second, short repetitive work cycle times have been shown to have a skewed distribution not dissimilar from the

gamma distribution. Rogers (134) reviewed the research of several masters degree candidates who studied work time distributions under the direction of Doctors R. N. Lehrer and J. J. Moder at the Georgia Institute of Technology. All of this research indicated a skewed distribution of work cycle times. If the value a_j' is taken as the sum of work cycle times constituting one span of work activity, it is logical to employ the gamma density to describe the distribution of a_j' , since gamma variables with the same basic parameter convolve into gamma variables. The third reason for selecting the gamma distribution is that the parameters may be conveniently altered by convolving additional exponential variables.

In the second case considered, the densities of a_j and a_j' are assumed to be truncated normal densities. The truncation is at zero in both cases. There are two reasons for selecting these densities as representative. First, if a_j and a_j' are each sums of numerous subspans, the central limit theorem may be employed. Second, the parameters in the densities may be altered with ease, and the effects of decreasing the coefficient of variation may readily be observed. Much of the applied work in the field of time study has been based on the assumption of normality of element times in work cycles, and normal variables convolve into normal variables. The assumptions of normality mentioned have been criticized severely (45), and experimental evidence has been presented by Davidson and others to contradict the assumption; however, the experimental evidence of particular situations does not preclude the existence of situations in which the

normality assumption is valid.

Consultation with mathematicians having knowledge of stochastic processes has indicated some difficulty associated with the analytical determination of the correlogram associated with the processes considered in this section. The function $R(u)$ defined in Equation (72) may be further analyzed as follows:

$$\begin{aligned} R(u) &= E\{X(t) X(t+u)\} = P\{X(t) = 1, X(t+u) = 1\} \quad (128) \\ &= P\{X(t) = 1\} P\{X(t+u) = 1 \mid X(t) = 1\}, \end{aligned}$$

and

$$\begin{aligned} P\{X(t+u) = 1 \mid X(t) = 1\} &= \int_0^\infty \left[\sum_{k=0}^\infty P\{X(t+u) = 1, \text{no. changes} \right. \\ &\quad \left. = 2k \mid X(t) = 1, y\} \right] f_1(y) dy, \end{aligned}$$

where the dummy variable y denotes the time that $X(t)$ has been equal to 1 at time t , and where $f_1(\cdot)$ is the density of a_j . The difficulties associated with the evaluation of this function led to the use of Monte Carlo techniques for the determination of the correlogram. Although it may be possible to evaluate Equation (128) in the case where the gamma densities have the same basic parameter, the analysis of the case where the parameters are different is prohibitively difficult. This entire question is a mathematical question, and the mathematician consulted had no knowledge of the existence of any literature in which this problem is treated. From an engineering standpoint, the use of Monte Carlo simulation yields adequate results; although the technique was

expensive in terms of the large amount of electronic computer time required.

Simulation Experiments

Gamma Distributions of the Span Lengths of A and A'. A period of 480 minutes was employed, and fifty realizations are simulated for this period. An electronic computer was employed, and the gamma variables were generated by convolving exponential variables. The densities of a_j and a_j' are

$$\begin{aligned} f_1(a_j) &= \frac{\lambda}{\Gamma(r)} (\lambda a_j)^{r-1} e^{-\lambda a_j} ; a_j \geq 0 \\ &= 0 ; \text{ otherwise} \end{aligned} \quad (129)$$

$$\begin{aligned} f_2(a_j') &= \frac{\lambda'}{\Gamma(r')} (\lambda' a_j')^{r'-1} e^{-\lambda' a_j'} ; a_j' \geq 0 \\ &= 0 ; \text{ otherwise} \end{aligned}$$

where r and r' exponential densities with parameters λ and λ' are convolved to obtain each of the a_j, a_j' ; $j = 1, 2, 3, \dots$. The generation of values for a_j and a_j' is in such a manner that a_j and a_j' are mutually independent random variables for $j = 1, 2, 3, \dots$.

The 480 minute period corresponds to one shift of continuous operation; however if the time dimension is altered to half minutes, quarter-minutes, etc., the results remain unaltered except for the scale change. Each 480 minute realization is generated independently, so that the activity state at the end of one realization does not affect the beginning state of the next realization. This was a matter of choice rather than necessity. Another procedure would have been to continuously

treat the $50 \times 480 = 24,000$ minutes of realization. The reason for selecting the method employed was that it was felt that many activities in the industrial community behave in such a manner that the idle or work state at the end of one shift does not affect the corresponding beginning state of the next shift. In order to simulate stability, the state A was the starting state with probability p , and A' was the starting state with probability $1 - p$.

The choice of values λ , λ' , r and r' was based on judgement; however the criteria were that the range of CV values between 0.10 and 1.5 should be represented, and several values for the ratio of the mean of a_j to the mean of a_j' ($\frac{r}{\lambda} / \frac{r'}{\lambda'}$) between 0.2 and 4 should be represented. These selections were such that it is felt that they are representative of the real world. A value $CV = 0.1$ implies that the interval,

$$E(Y_j) \pm 0.3 E(Y_j) = E(Y_j) [1.0 \pm 0.3] ,$$

contains most of the distribution of Y_j . Conversations with work study practitioners indicate that this represents a "regular" pattern of activity behavior. A CV value of 1.5 indicates that most of the distribution of Y_j is contained on the interval $[0, 5.5 E(Y_j)]$, and this is considered a very "irregular" pattern of activity behavior. The values ($\frac{r}{\lambda} / \frac{r'}{\lambda'}$) were selected to satisfy two criteria. First, several of the ratios ($\frac{r}{\lambda} / \frac{r'}{\lambda'}$) were held constant in order to observe the effect of decreasing CV, and second, some differing values of ($\frac{r}{\lambda} / \frac{r'}{\lambda'}$) were selected to observe the effect of varying this ratio. It was felt that values 0.2 to 4.0 are not uncommon in practice; therefore several values

in the range were employed. It was necessary to limit the sets of parameters to 10 due to the large amount of electronic computer time required.

For each realization, the values of the random variables a_j and a'_j are calculated as follows:

1. For a_j , r values of independent, exponential random variables, each with parameter λ , are added. The values for the exponential random variables were obtained by the following relation, where RN'_i is the i^{th} random number generated by the multiplicative-congruential method (33).

$$\text{The } i^{\text{th}} \text{ value of the exponential random variable} = \left\lfloor \frac{\ln(RN'_i)}{\lambda} \right\rfloor \quad (130)$$

2. For a'_j , r' values of independent, exponential random variables, each with parameter λ' , are added. The values for the exponential random variables are obtained in the same manner as above with λ' replacing λ . The value of i is increased before each usage of the relation for calculating values of the exponential random variables.

3. Steps one and two are repeated until ϕ_j as defined in Equation (105) is at least 480.

4. After ϕ_j is sufficiently large to cover the 480 minute realization the random number generator is again employed to order the occurrence of A and A' , that is, the probability that the realization starts in activity A is p .

5. At this point the realization is observed at 30 second intervals, and the results as 0's and 1's are sequentially punched into cards with 30 observations per card and 32 cards per realization. The 50 realizations produce a box of 1,600 cards.

The input for the random number generator is a ten digit integer constant, C_0 , and an initial ten digit random integer, RN . The congruential method employed consists of the following two steps:

1. $RN_i = C_0(RN_{i-1}) \mod 10^{10}$
2. $RN'_i = 10^{-10}(RN_i)$, where RN_i is a ten digit integer.

Each ten digit integer, RN_i , depends only on the previous integer, and this integer is always preserved for the next use of the generator.

The electronic computer program for the generation of process realizations is given in Appendix B, Table 12. The total time required for generating 50 realizations for a particular set of parameters varied from 25 to 104 minutes depending upon the values r and r' . The values of the parameters employed are given in Table 3.

Normal Distributions of the Span Lengths of A and A' . A period of 480 minutes was again employed, and the normal variables truncated at zero were generated by convolving ten uniform variables and discarding all values less than zero. The densities of a_j and a'_j are approximately as follows:

$$\begin{aligned}
 f_1(a_j) &= \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{a_j - \mu}{\sigma}\right)^2}}{\int_0^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{t - \mu}{\sigma}\right)^2} dt} ; \quad a_j \geq 0 \\
 &= 0 ; \text{ otherwise} \\
 f_2(a'_j) &= \frac{\frac{1}{\sigma'\sqrt{2\pi}} e^{-1/2\left(\frac{a'_j - \mu'}{\sigma'}\right)^2}}{\int_0^{\infty} \frac{1}{\sigma'\sqrt{2\pi}} e^{-1/2\left(\frac{t - \mu'}{\sigma'}\right)^2} dt} ; \quad a'_j \geq 0 \\
 &= 0 ; \text{ otherwise,}
 \end{aligned} \tag{131}$$

Table 3. Values of Parameters Employed in Simulation with Gamma Distributions for Lengths of Span A and A'

| λ | r | Mean r/λ | λ' | r' | Mean $\frac{r'}{\lambda'}$ | $\frac{r}{\lambda} / \frac{r'}{\lambda'}$ | Mean Cycle Length $\frac{r'}{\lambda'} + \frac{r}{\lambda}$ | Cycle Length Variance $\frac{r}{(\lambda)^2} + \frac{r}{(\lambda')^2}$ | Coeffi- cient of Variation of Cycle Length |
|-----------|-----|---------------------|------------|------|-------------------------------|---|--|---|--|
| 0.40 | 2 | 5 | 0.08 | 2 | 25 | 0.2 | 30.0 | 2512.500 | 1.667 |
| 0.10 | 1 | 10 | 0.10 | 1 | 10 | 1.0 | 20.0 | 200.000 | 0.707 |
| 1.00 | 10 | 10 | 1.00 | 10 | 10 | 1.0 | 20.0 | 20.000 | 0.224 |
| 3.00 | 9 | 3 | 2.00 | 14 | 7 | 0.43 | 10.0 | 4.500 | 0.212 |
| 4.00 | 16 | 4 | 4.00 | 16 | 4 | 1.0 | 8.0 | 2.000 | 0.177 |
| 2.00 | 20 | 10 | 2.00 | 20 | 10 | 1.0 | 20.0 | 10.000 | 0.150 |
| 5.00 | 30 | 6 | 4.00 | 16 | 4 | 1.5 | 10.0 | 2.200 | 0.149 |
| 4.00 | 64 | 16 | 4.00 | 16 | 4 | 4.0 | 20.0 | 5.000 | 0.112 |
| 10.00 | 50 | 5 | 10.00 | 30 | 3 | 1.67 | 8.0 | 0.800 | 0.112 |
| 5.00 | 50 | 10 | 5.00 | 50 | 10 | 1.0 | 20.0 | 4.000 | 0.100 |

where a_j and a_j' are mutually independent random variables for $j = 1, 2, 3, \dots$.

The choice of values μ , μ' , σ , and σ' was based on judgment as in the case of the gamma densities. The criteria were the same, except that one very "regular" pattern in which $CV = 0.003$ was investigated; and one case was investigated in which $\sigma' = 0$, so that a_j' was a constant.

For each simulated realization, the values of a_j and a_j' are calculated as follows:

1. For a_j , ten values of independent random numbers from a uniform distribution on $[0,1]$ are added, and

$$a_j = \sigma \left[\frac{\left(\sum_{i=1}^{10} RN_i' \right) - 5.0}{\sqrt{5/6}} \right] + \mu \quad (132)$$

The congruential-multiplicative method of random number generation is employed. The value of a_j is tested, and if $a_j < 0$, it is discarded, and the above procedure is repeated.

2. For a_j' , ten values of independent random numbers from a uniform distribution on $[0,1]$ are added, and

$$a_j' = \sigma' \left[\frac{\left(\sum_{i=1}^{10} RN_i' \right) - 5.0}{\sqrt{5/6}} \right] + \mu' \quad (133)$$

The congruential-multiplicative method of random number generation is again employed, and a_j' is again tested against zero. If $a_j' < 0$, it is discarded, and the above procedure is repeated.

3. After steps one and two are completed for a particular j , ϕ_j as defined in Equation (105) is tested against 480, and steps one and two are repeated until ϕ_j is at least 480.

4. When ϕ_j is sufficiently large to cover the 480 minute realization, the random number generator is again employed to start the realization in state A with probability p .

The realization is observed at 30 second intervals, and the results as 0's and 1's are sequentially punched into cards with 30

observations per card and 32 cards per realization. The electronic computer program for the generation of process realizations is given in Appendix B, Table 13. The total time required for the generation of 50 process realizations was 49 minutes. The values of the parameters employed are given in Table 4.

Table 4. Values of Parameters Employed in Simulation with Truncated Normal Distributions for Lengths of Span A and A'

| μ | μ' | μ/μ' | σ^2 | σ'^2 | Approx. Mean Cycle Length | Approx. Cycle Length Variance | Approx. Coeffi- cient of Variation |
|-------|--------|------------|------------|-------------|------------------------------------|--|---|
| 5.0 | 5.0 | 1.0 | 6.0 | 6.0 | 10.0 | 12.0 | 0.347 |
| 10.0 | 10.0 | 1.0 | 6.0 | 6.0 | 20.0 | 12.0 | 0.173 |
| 10.0 | 10.0 | 1.0 | 4.0 | 4.0 | 20.0 | 8.0 | 0.141 |
| 10.0 | 10.0 | 1.0 | 2.0 | 2.0 | 20.0 | 4.0 | 0.100 |
| 4.0 | 20.0 | 0.2 | 1.0 | 4.0 | 24.0 | 5.0 | 0.093 |
| 30.0 | 5.0 | 6.0 | 4.0 | 4.0 | 35.0 | 8.0 | 0.081 |
| 10.0 | 10.0 | 1.0 | 2.0 | 0.0 | 20.0 | 2.0 | 0.071 |
| 5.0 | 30.0 | 0.167 | 1.0 | 4.0 | 35.0 | 5.0 | 0.064 |
| 4.0 | 20.0 | 0.2 | 0.25 | 0.25 | 24.0 | 0.5 | 0.003 |

Calculation of Serial Correlations

Formulation. Equation (75) is utilized for the calculation of an empirical correlation function from each simulated process realization. The symbol $\rho_{u,i}$ represents the serial correlation between instants u minutes apart for the i^{th} realization, and the mean and standard deviation over realizations, i , are calculated as follows:

$$\rho_u = \frac{1}{50} \sum_{i=1}^{50} \rho_{u,i} , \quad (134)$$

and

$$s_{\rho_u} = \frac{1}{49} \sum_{i=1}^{50} (\rho_{u,i} - \rho_u)^2 \quad (135)$$

The computer program for calculating the quantities defined in Equations (134) and (135) is given in Appendix B, Table 14. The computer time required for each set of parameters was 109 minutes, and the cards from the simulation experiments served as input for this program.

Test of Logic and Computer Program for Simulation and Calculation of Serial Correlations. The test of computer programs and logic was conducted by employing the gamma densities with $\lambda = \lambda' = 0.1$ and $r = r' = 1$, so that the densities of a_j and a_j' reduce to exponential densities for which the theoretical correlogram is known. Figure 6 shows the values ρ_u for several values of RN and C_0 as well as the theoretical values and values obtained from employing random number input obtained from tables (130) rather than from internal generation in the simulation program. The digits in the tables were generated from an "electronic roulette" device giving "true" random numbers. A computer program differing from the program in which the random numbers were internally generated was required in this last case, and this program is also given in Appendix B, Table 15. Tabular results corresponding to Figure 6 are given in Tables 17, 18, 19, and 20, of Appendix B.

Visual analysis of Figure 6 led to the selection of $C_0 = 5634765623$ and $RN = 4999465853$ for the random number generator. This selection was made due to the lack of periodicity in the correlogram and the general good behavior of the correlogram. No random generator of the type employed herein gives truly random numbers; however, it is not feasible to read random number input into the computer when r and r' are large.

Results

Preliminary Analysis. The resulting correlograms for the gamma densities are given in Figures 7 - 15, and the correlograms for the normal densities are given in Figures 16 - 24. Tabular results corresponding to Equations (134) and (135) are given in Tables 21 - 29 and Tables 30 - 38 of Appendix B for the gamma and truncated normal densities respectively.

Inspection of Figures 7 - 24 reveals that the correlogram is not monotone decreasing, convex for any of the parameters employed. Figure 7 displays the correlogram for the gamma densities comprising a cycle having a very high coefficient of variation, and the correlogram in this case comes closest to having the type form required for the application of Cochran's theorem.

The general behavior of the correlograms for all cases may be termed as damped periodic, and in the cases where a_j and a_j' have equal mean values, the fluctuations of ρ_u about zero are more symmetric for both gamma and truncated normal densities; however, this symmetry is not destroyed when

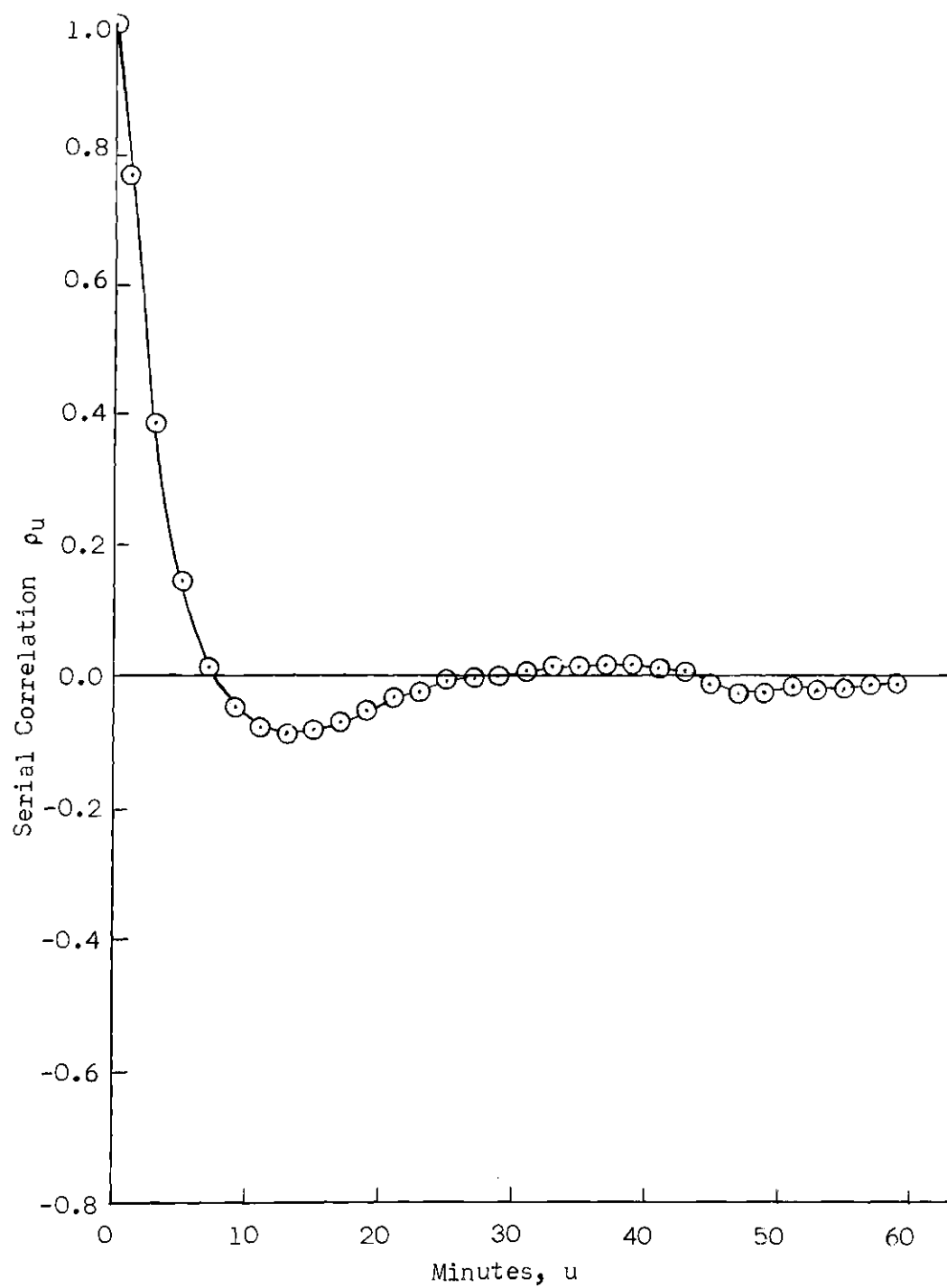


Figure 7. Correlogram from Simulation;
 $\lambda = 0.08$, $\lambda' = 0.4$, $r = r' = z$

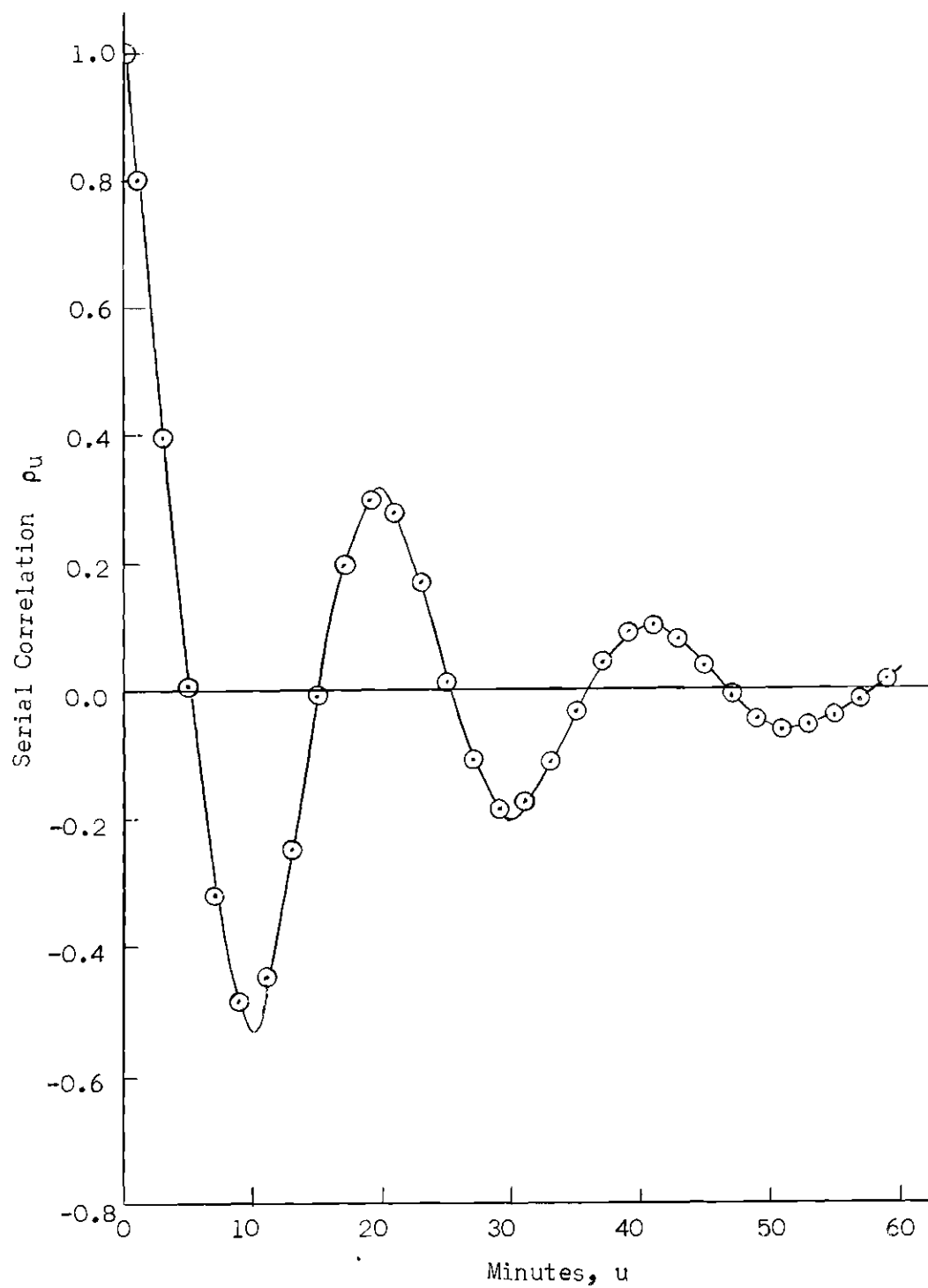


Figure 8. Correlogram from Simulation;
 $\lambda = \lambda' = 1.0$, $r = r' = 10$

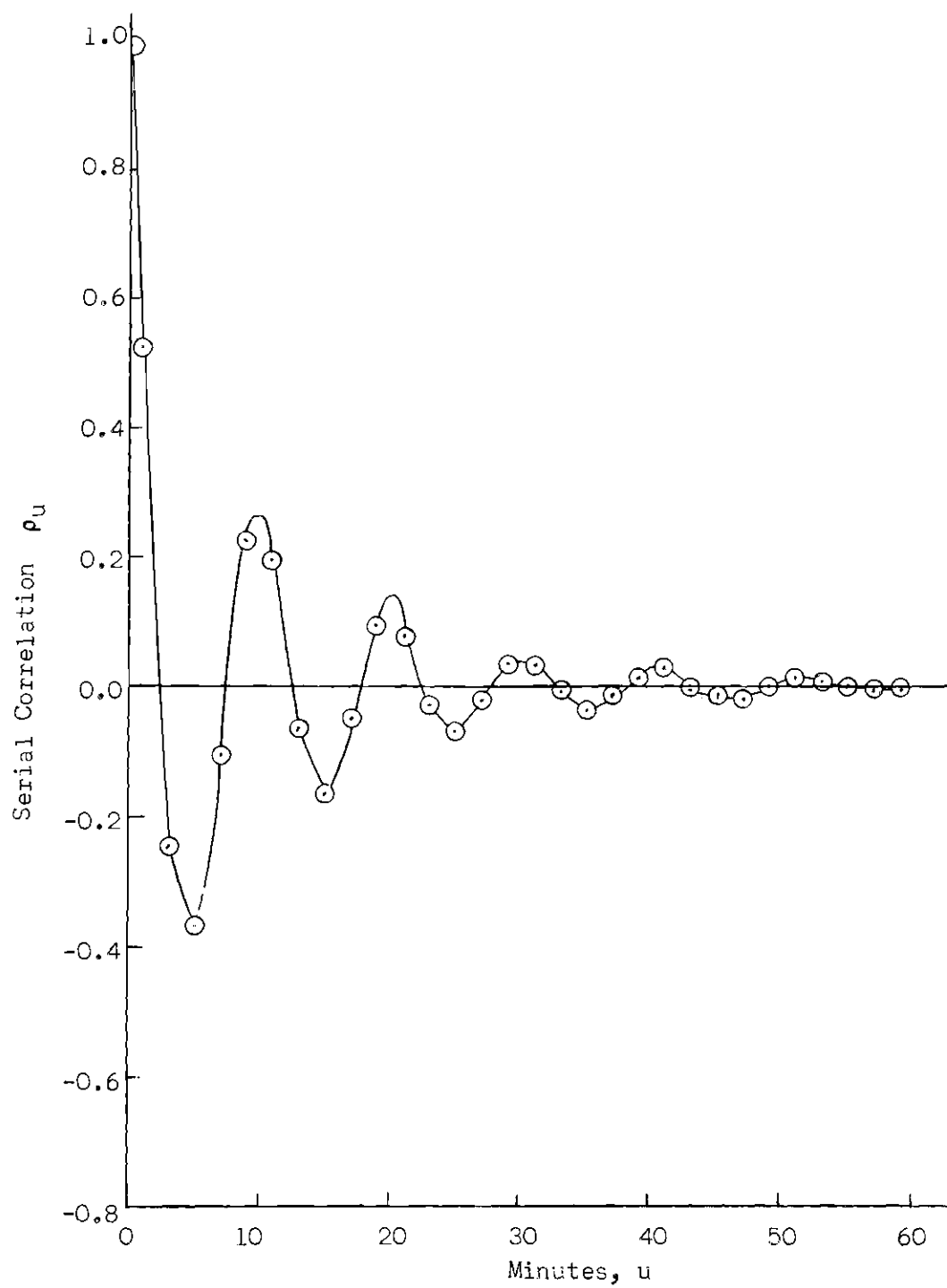


Figure 9. Correlogram from Simulation;
 $\lambda = 3.0$, $\lambda' = 2.0$, $r = 9$, $r' = 14$

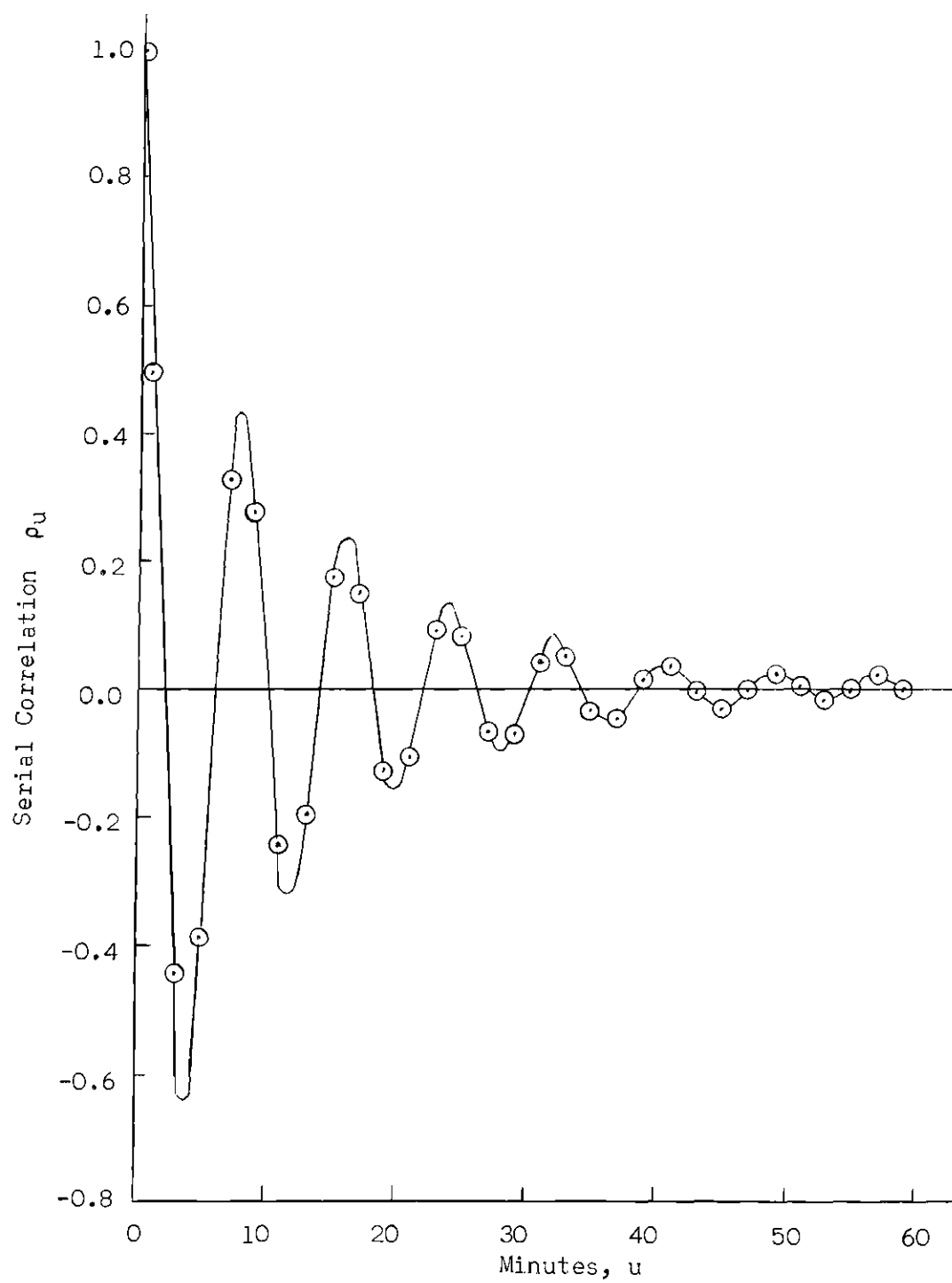


Figure 10. Correlogram from Simulation;
 $\lambda = \lambda' = 4.0$, $r = r' = 16$

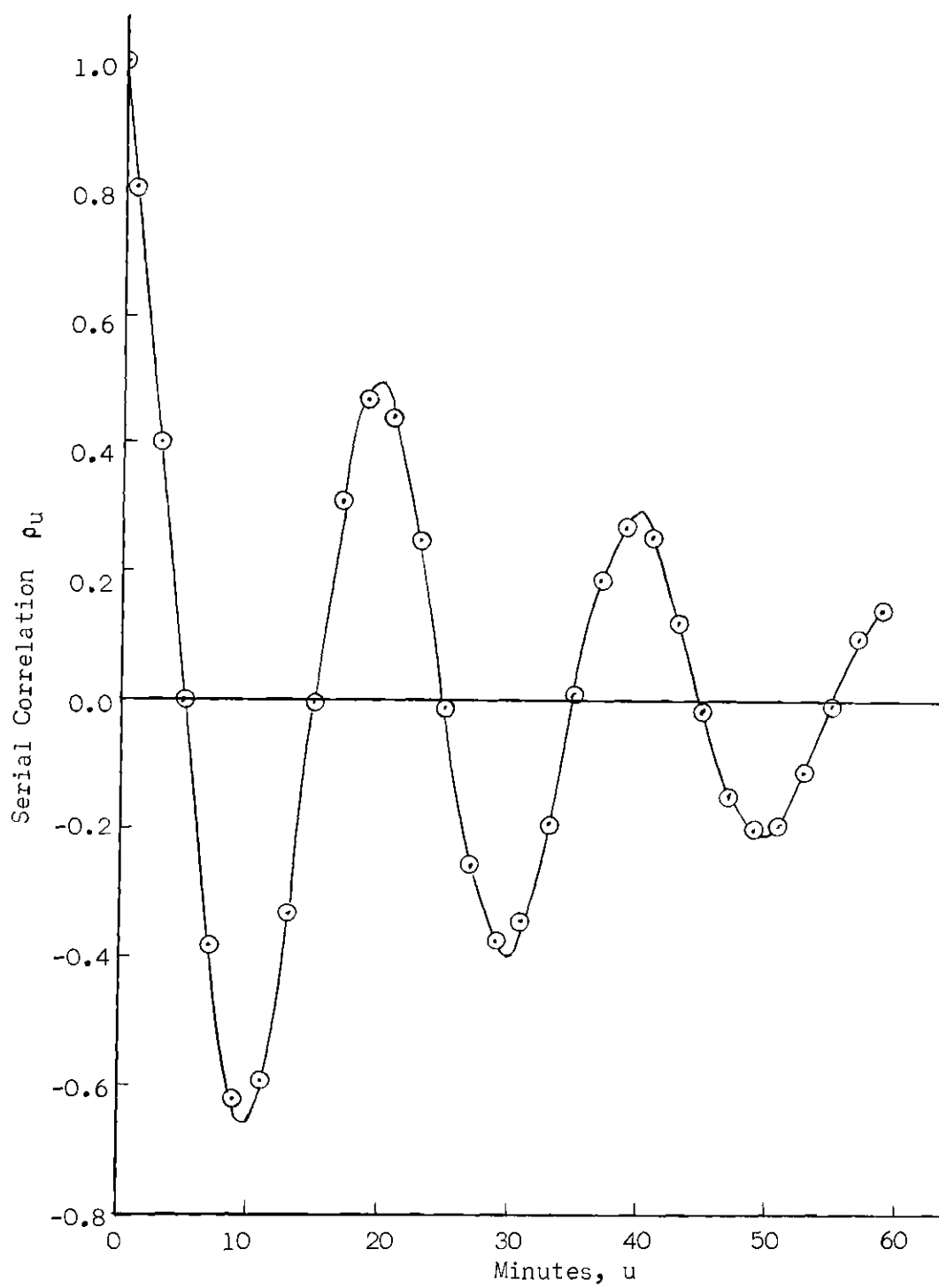


Figure 11. Correlogram from Simulation;
 $\lambda = \lambda' = 2.0$, $r = r' = 20$

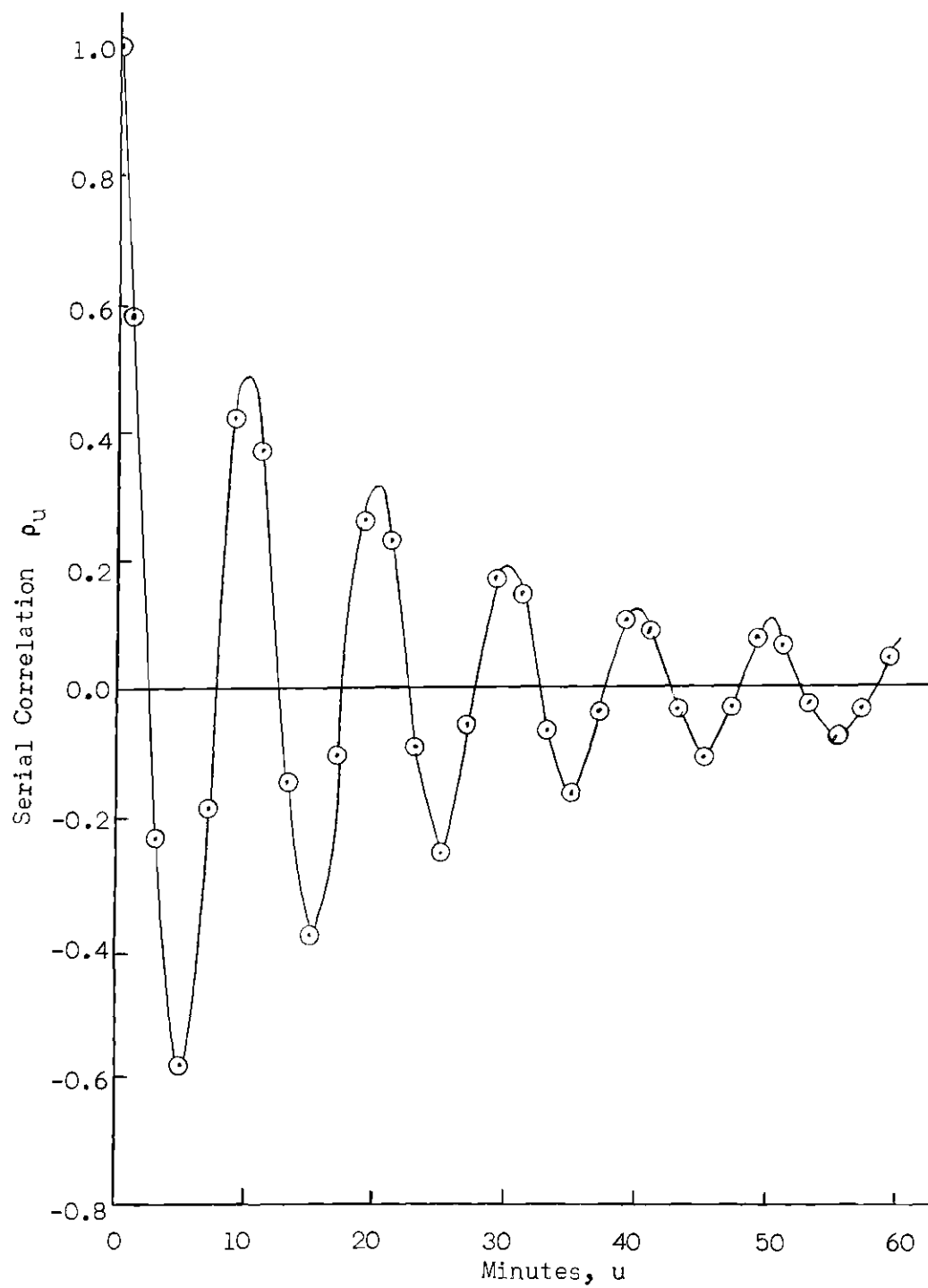


Figure 12. Correlogram from Simulation;
 $\lambda = 5.0$, $\lambda' = 4.0$, $r = 30$, $r' = 16$

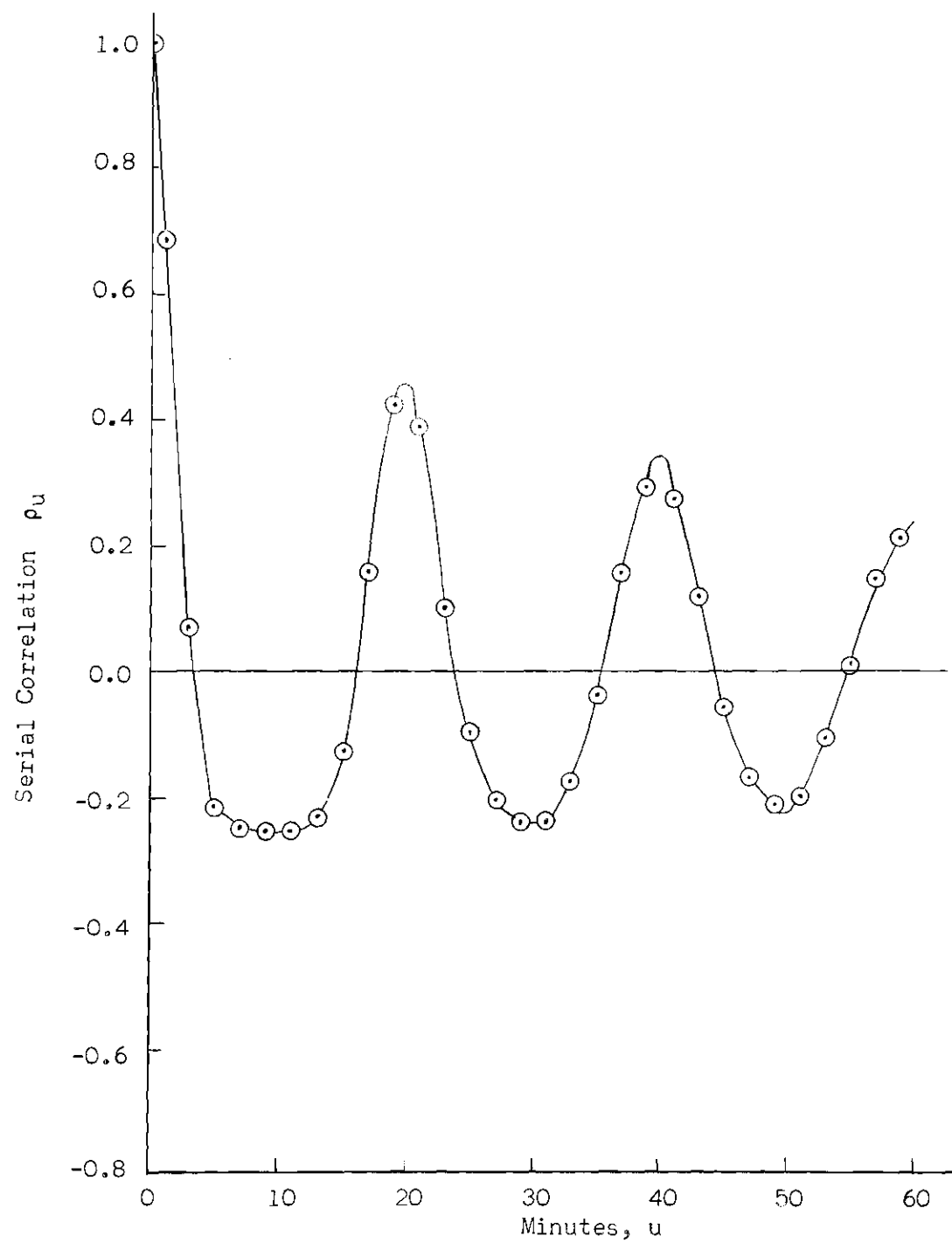


Figure 13. Correlogram from Simulation;
 $\lambda = 4.0$, $\lambda' = 4.0$, $r = 64$, $r' = 16$

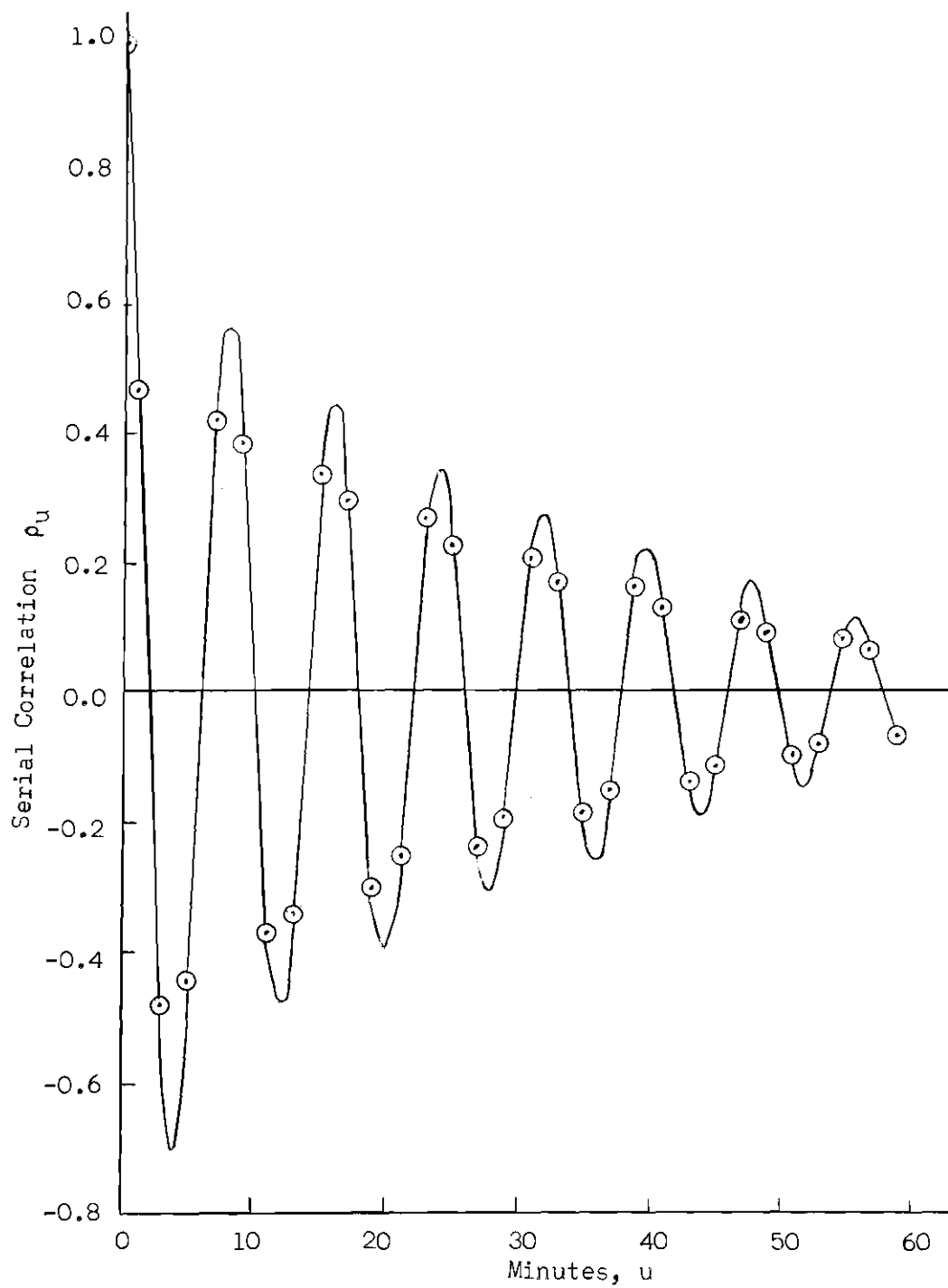


Figure 14. Correlogram from Simulation;
 $\lambda = \lambda' = 10.0$, $r = 50$, $r' = 30$

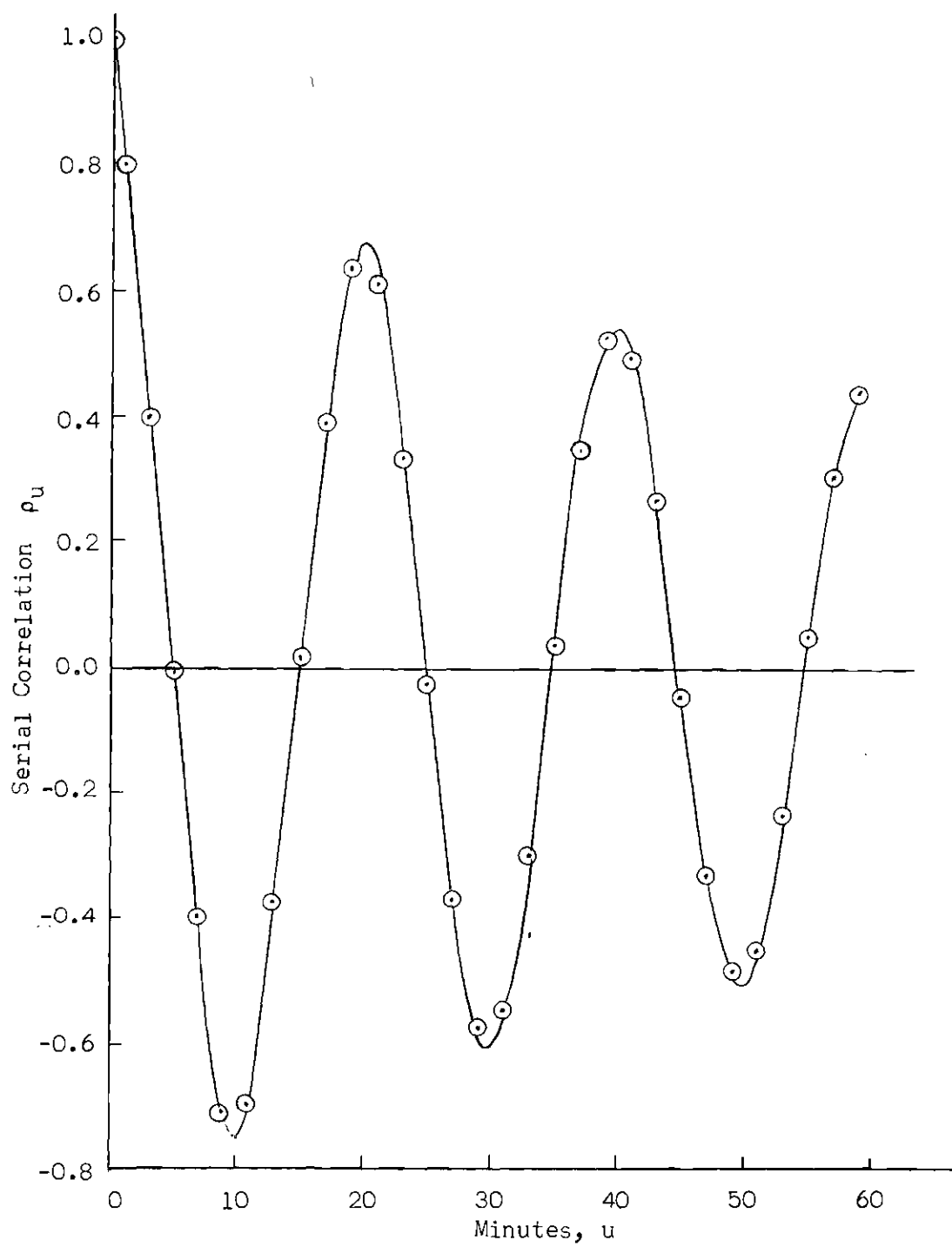


Figure 15. Correlogram for Simulation;
 $\lambda = \lambda' = 5.0$, $r = r' = 50$

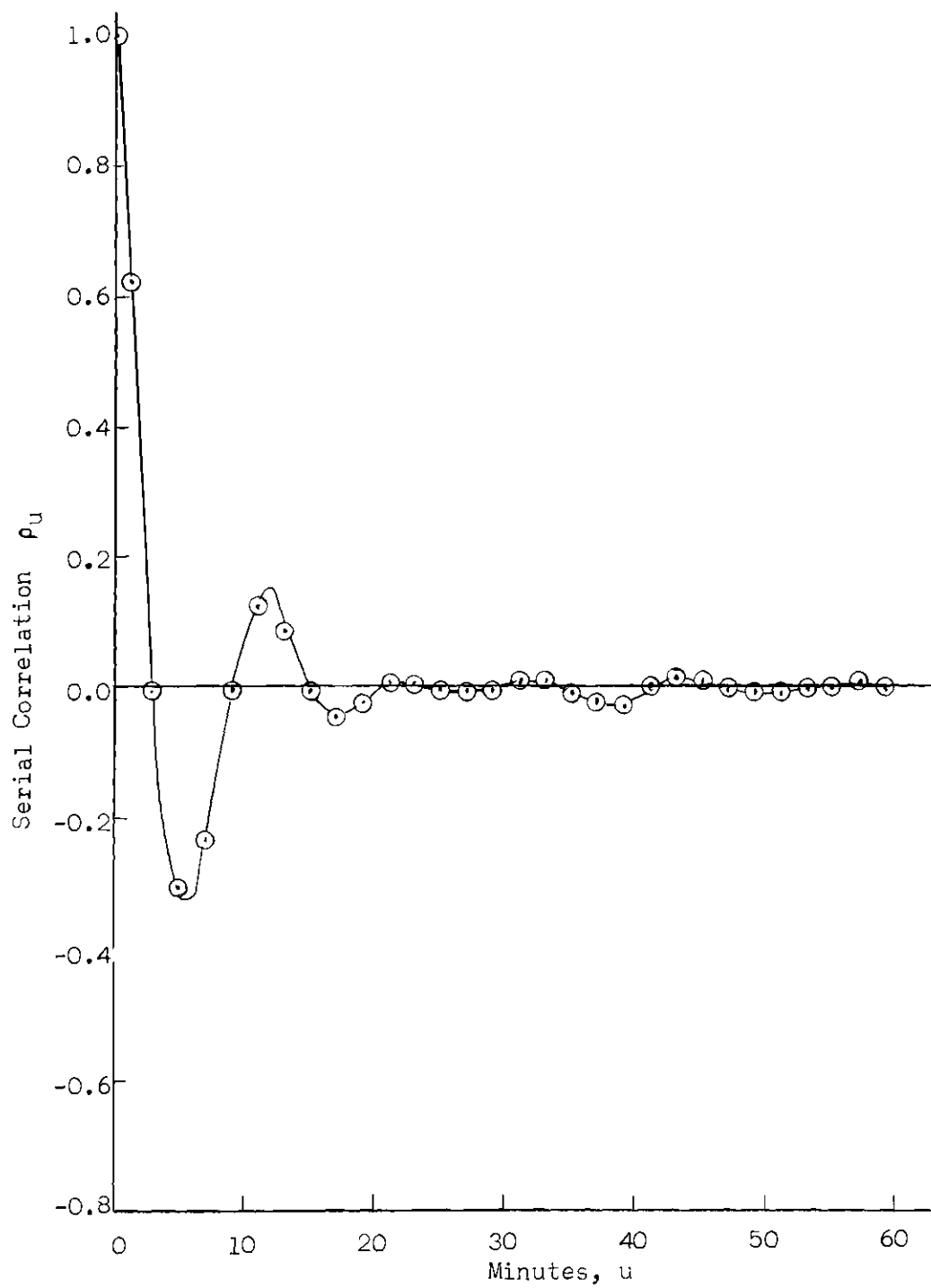


Figure 16. Correlogram from Simulation;
 $\mu = \mu' = 5.0$, $\sigma^2 = \sigma'^2 = 6.0$

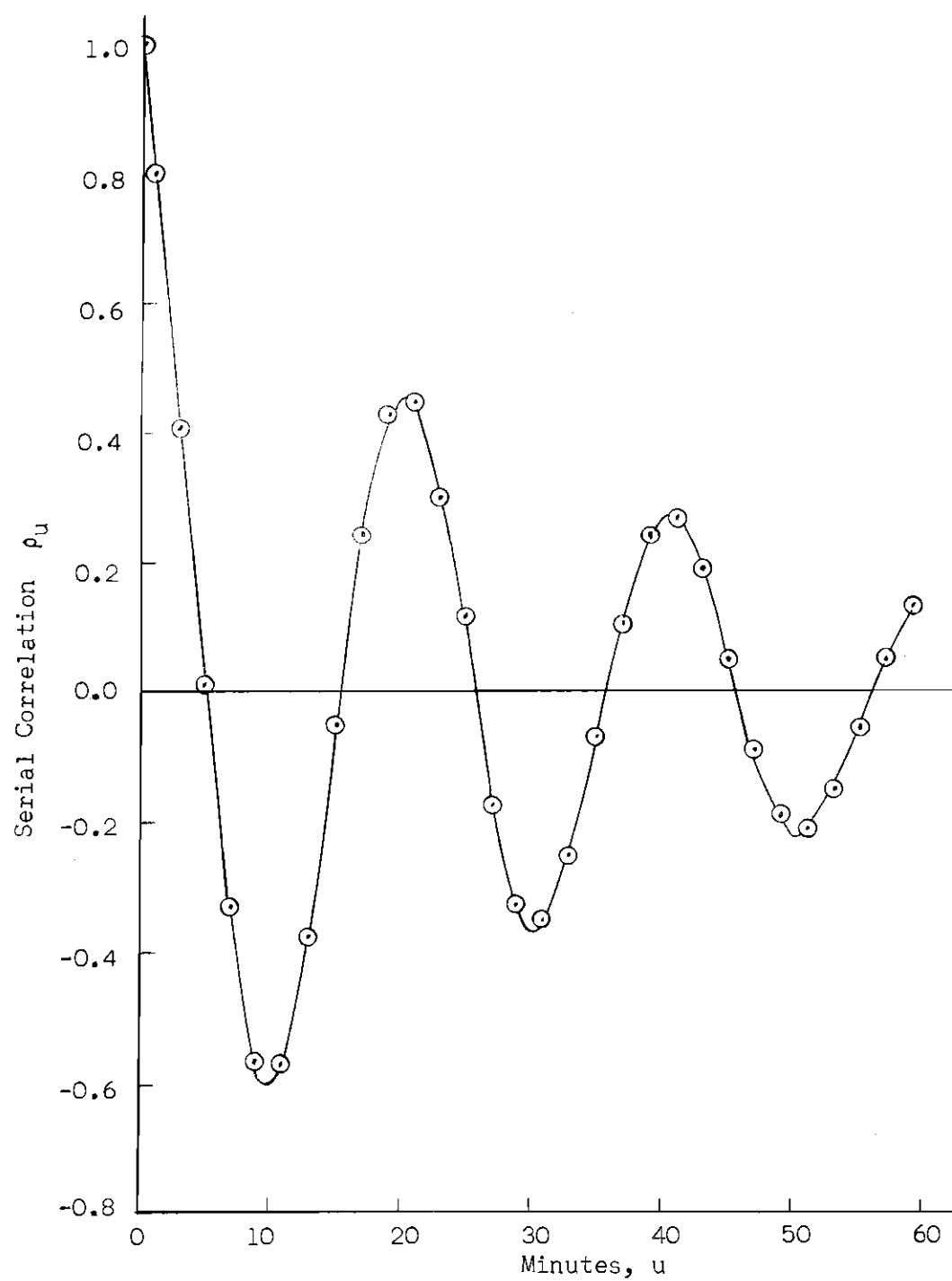


Figure 17. Correlogram from Simulation;
 $\mu = \mu' = 10.0$; $\sigma^2 = \sigma'^2 = 6.0$

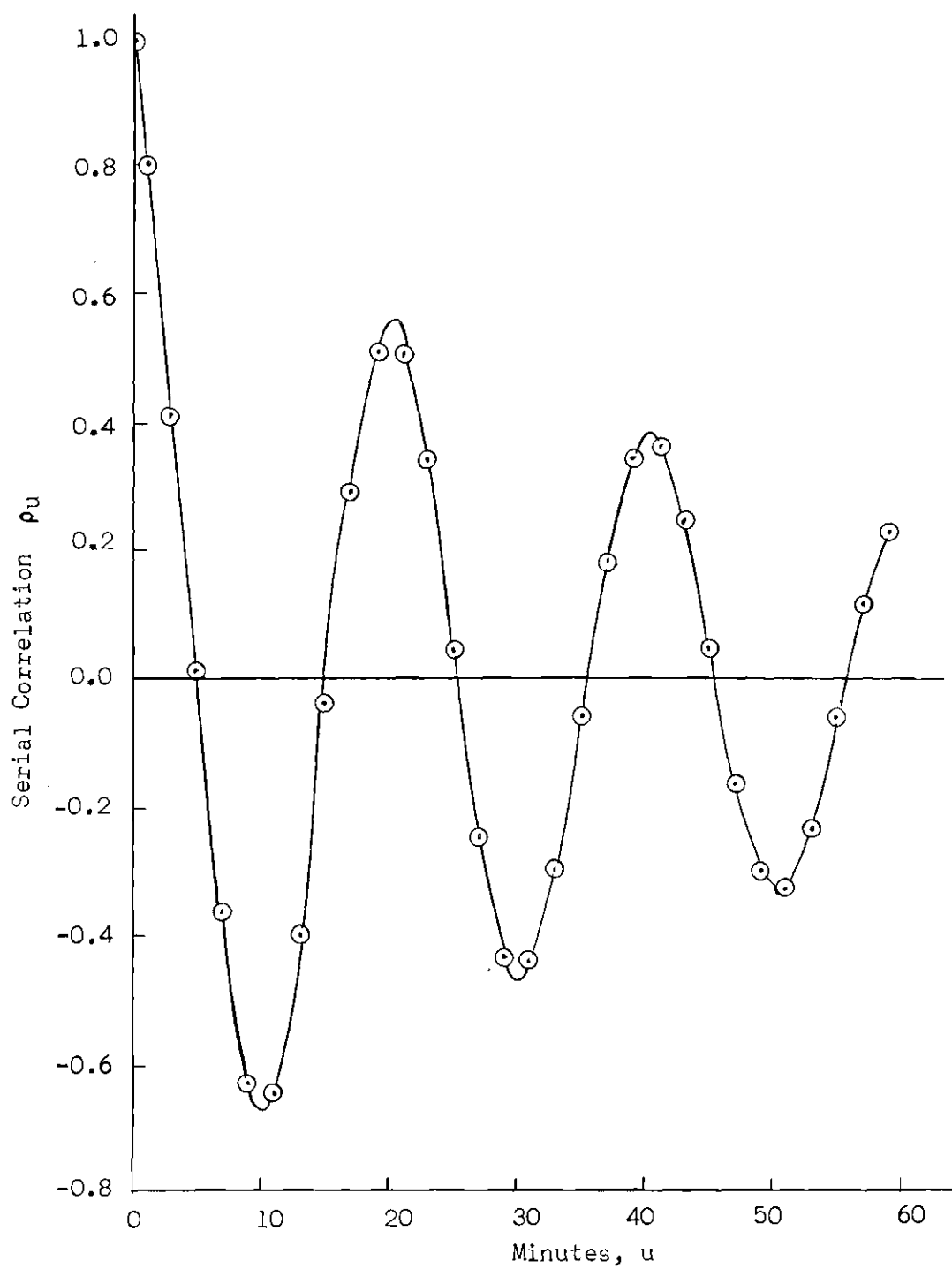


Figure 18. Correlogram from Simulation;
 $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 4.0$

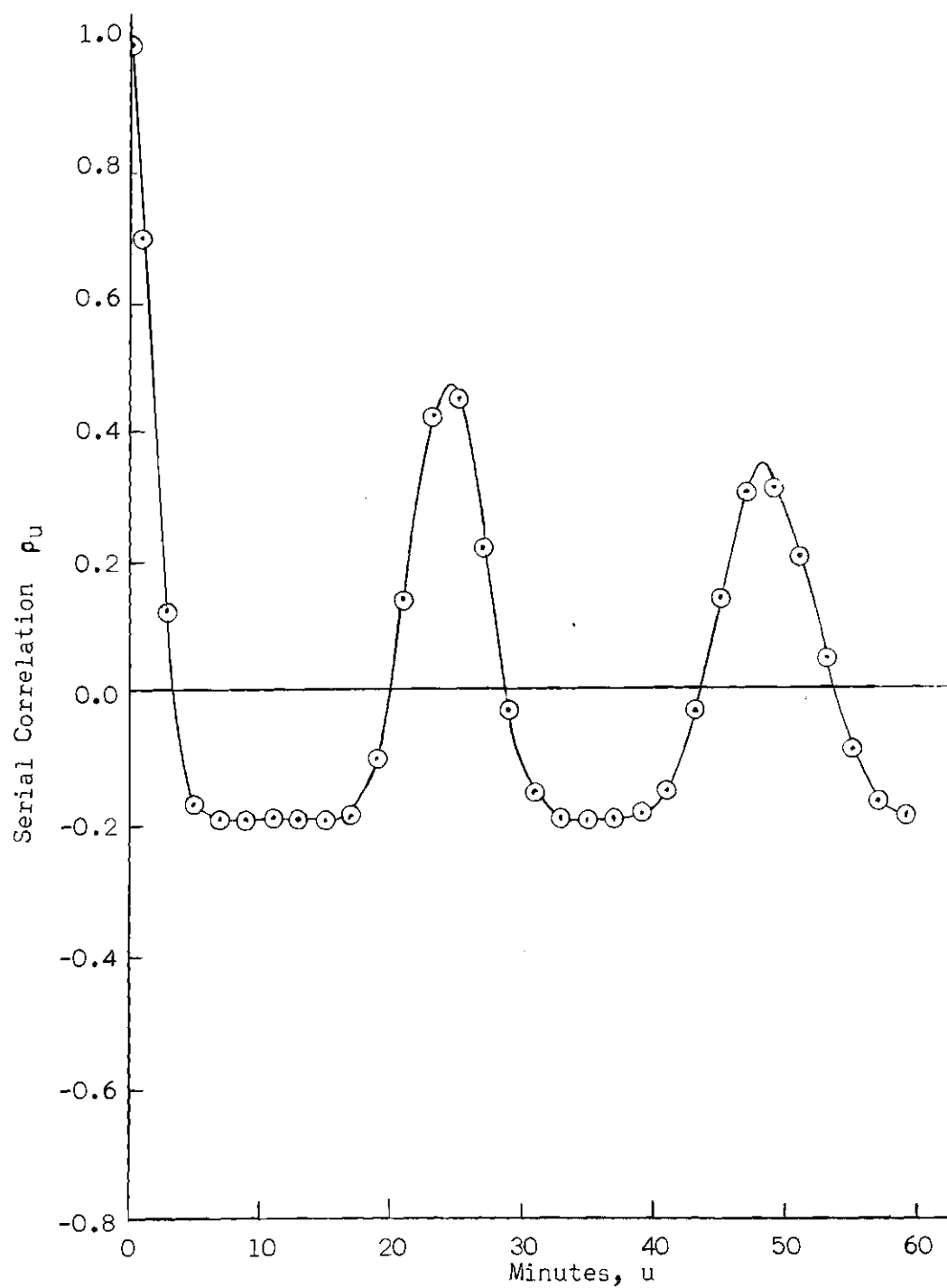


Figure 20. Correlogram from Simulation; $\mu = 4.0$,
 $\mu' = 20.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 4.0$

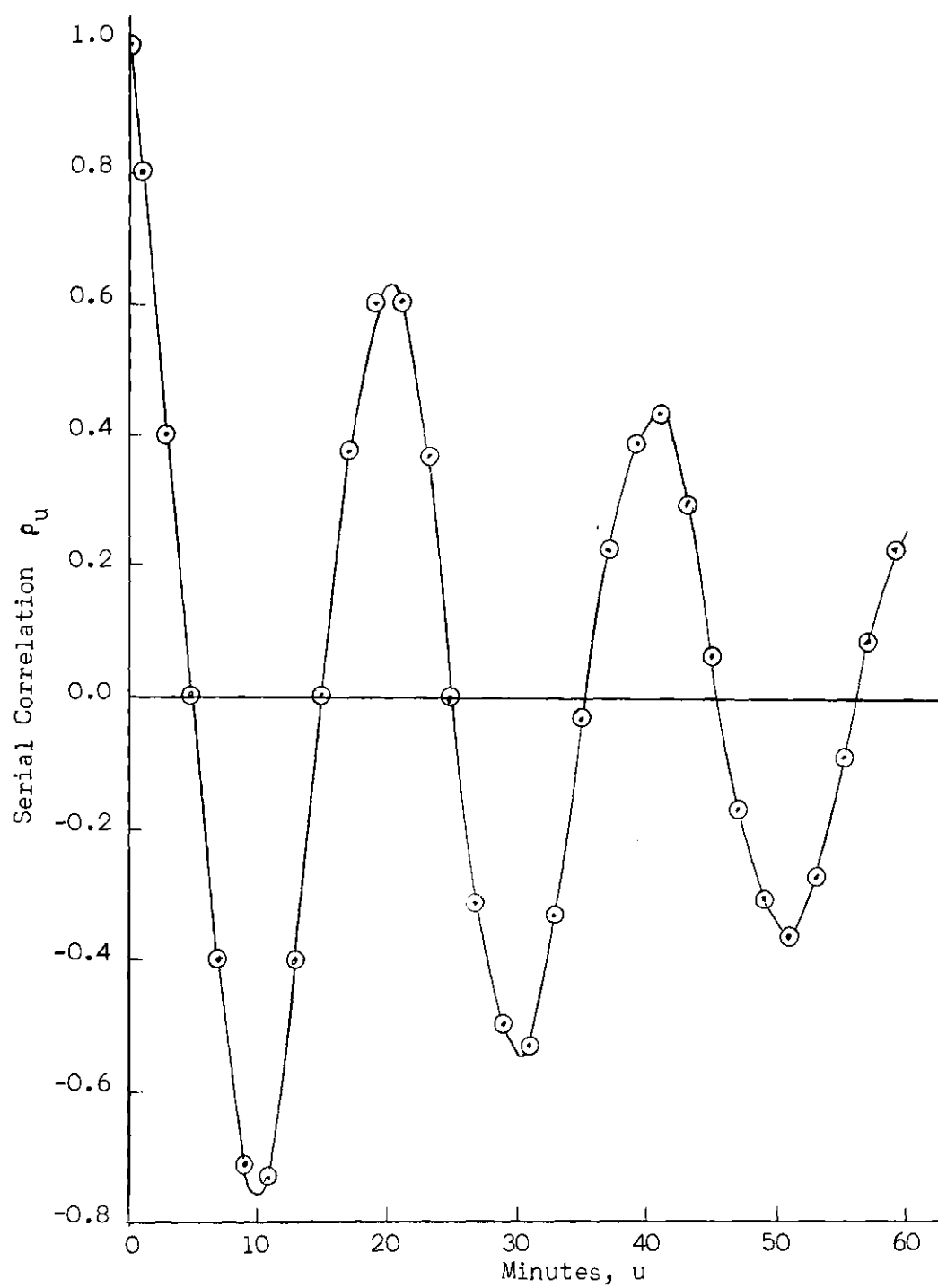


Figure 19. Correlogram from Simulation;
 $\mu = \mu' = 10.0$; $\sigma^2 = \sigma'^2 = 2.0$

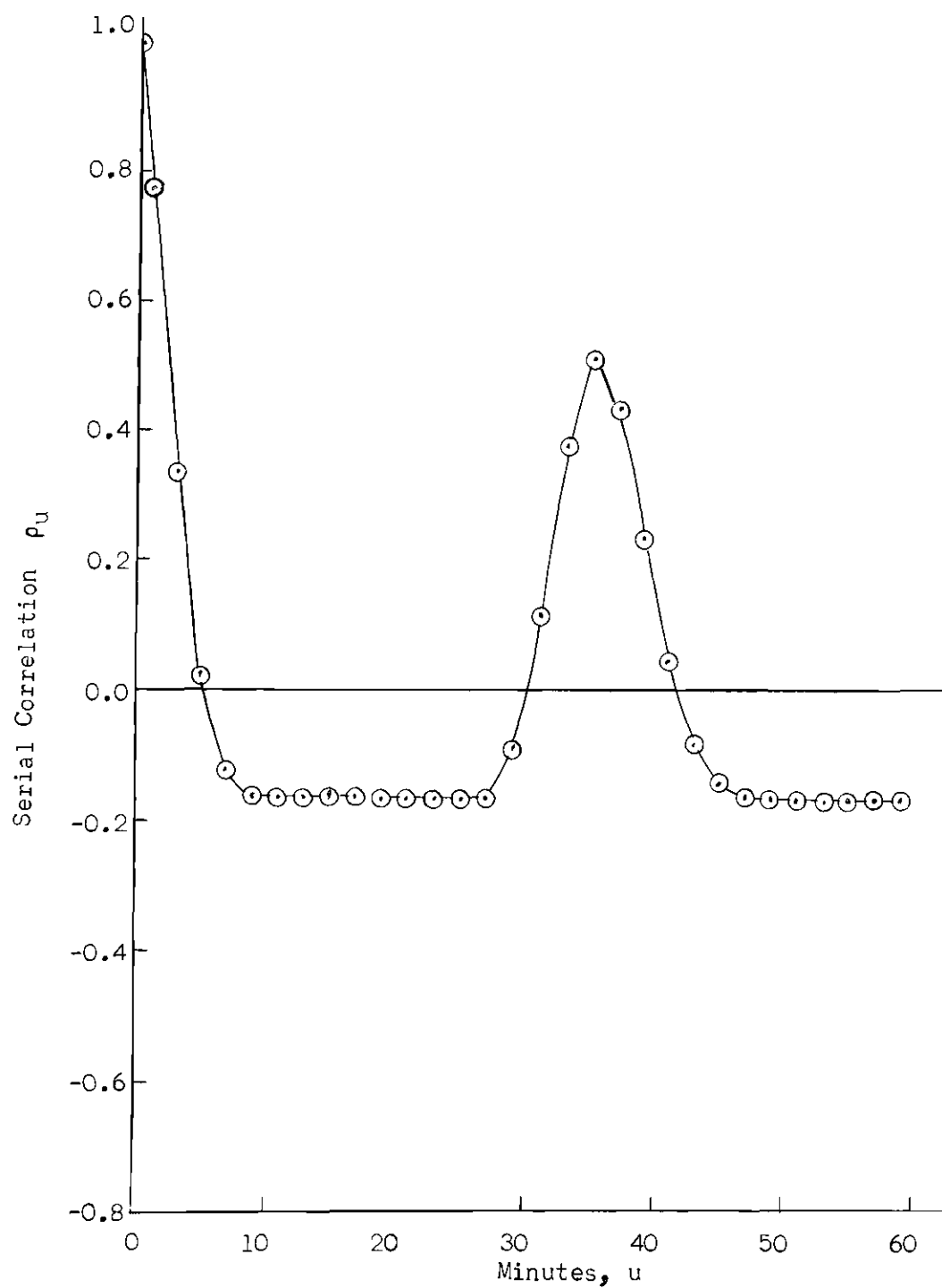


Figure 21. Correlogram from Simulation;
 $\mu = 30.0$, $\mu' = 5.0$, $\sigma^2 = \sigma'^2 = 4.0$

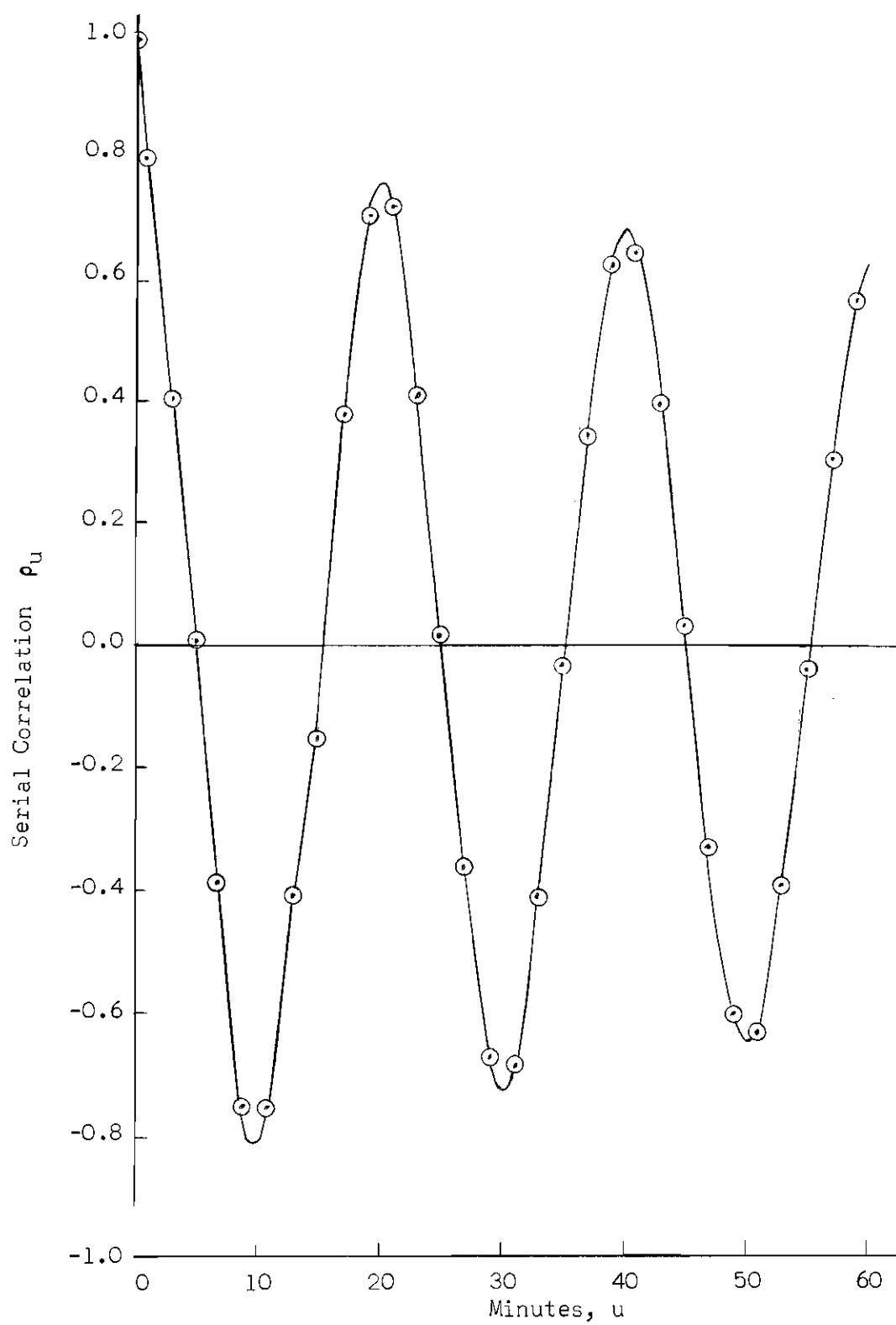


Figure 22. Correlogram from Simulation;
 $\mu = \mu' = 10.0$, $\sigma^2 = 2.0$, $\sigma'^2 = 0.0$

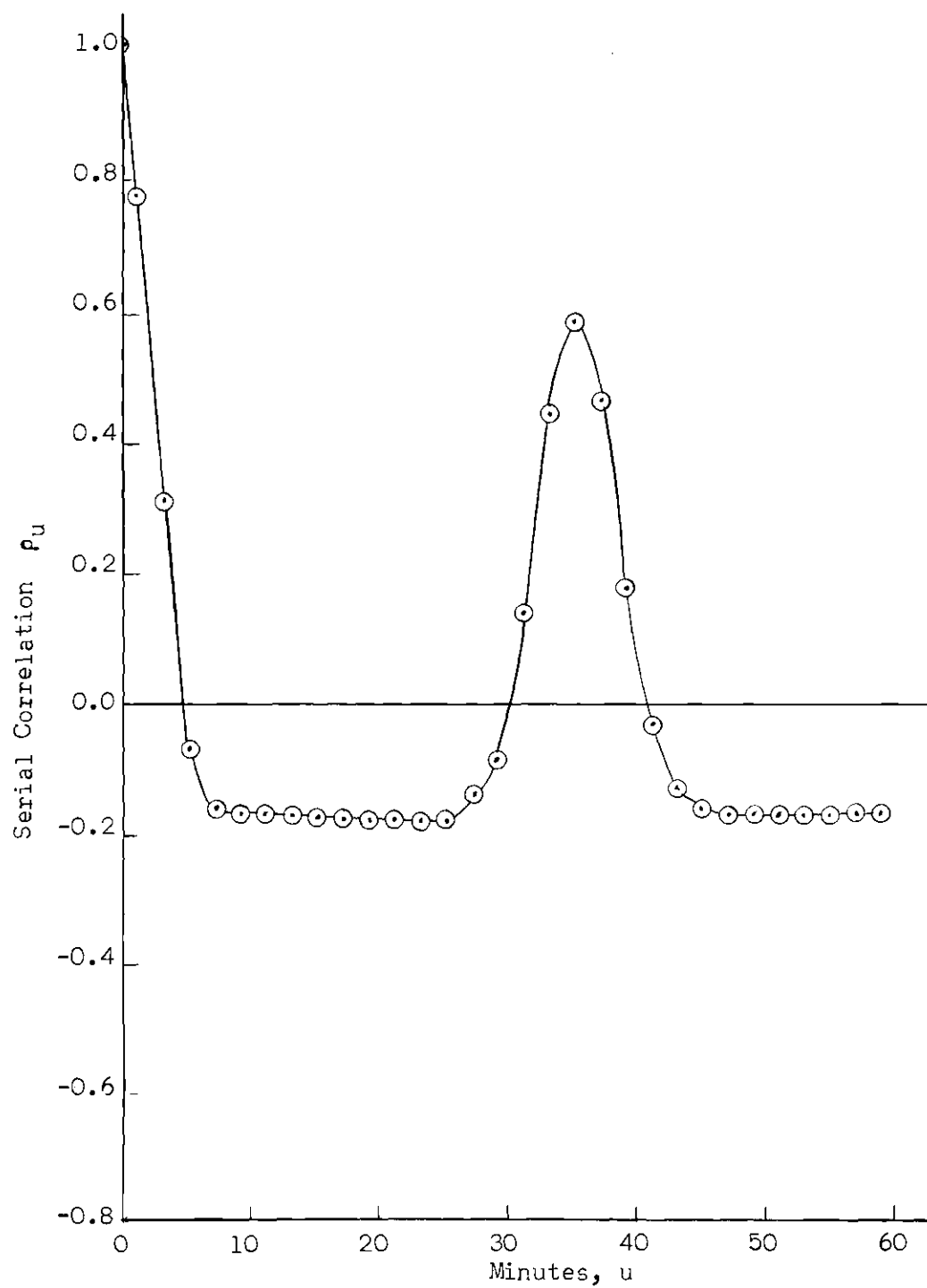


Figure 23. Correlogram from Simulation;
 $\mu = 5.0$, $\mu' = 30.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 5.0$

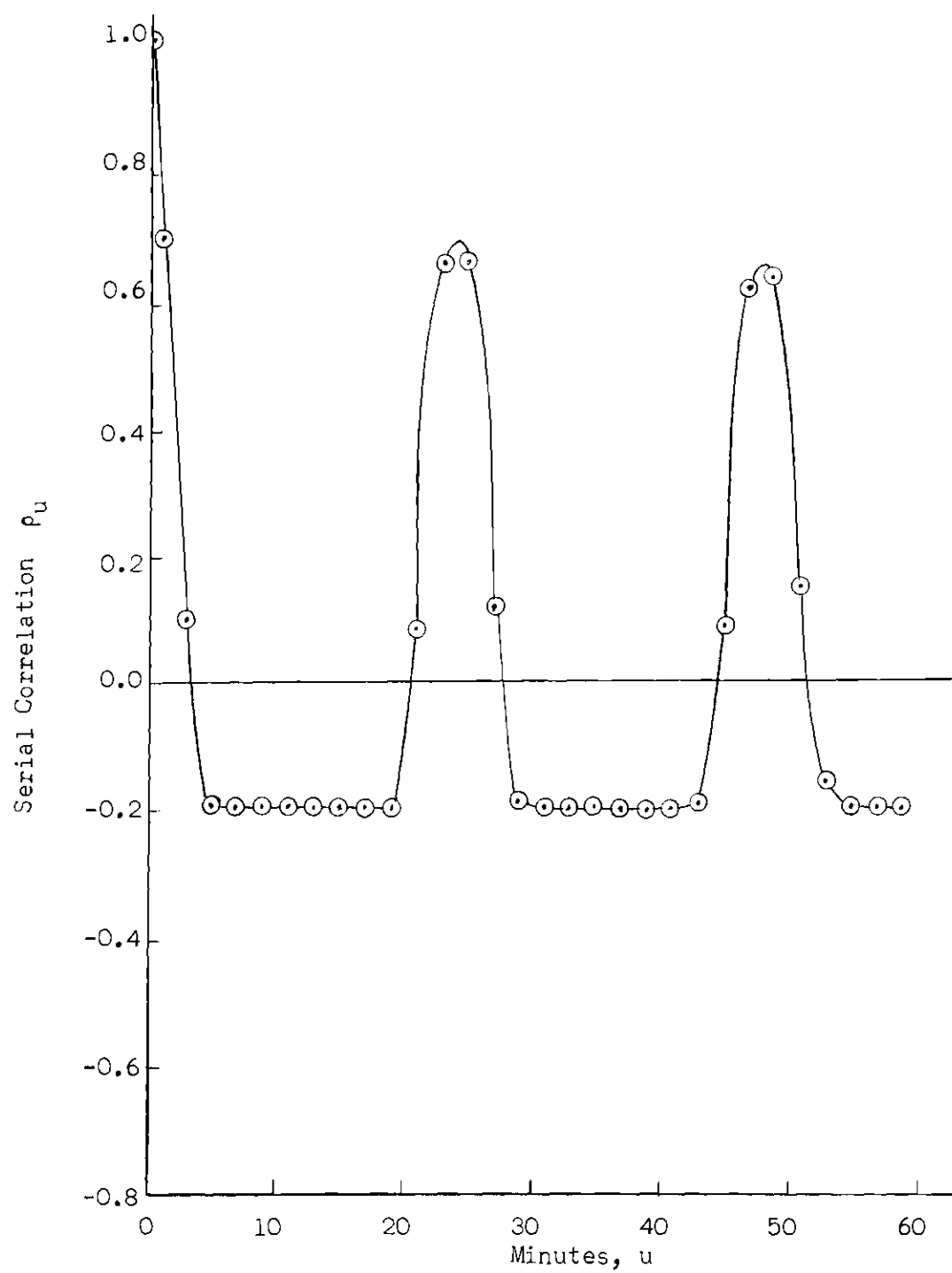


Figure 24. Correlogram from Simulation; $\mu = 5.0$,
 $\mu' = 30.0$, $\sigma^2 = 1.0$, $\sigma'^2 = 5.0$

$$C = \frac{\text{Max } [E(a_j), E(a_j')] }{\text{Min } [E(a_j), E(a_j')] } < 3 .$$

This fact is observed in Figures 10, 13, and 15. Figures 14, 21, 22, 24, and 25 correspond to cases where $C = 4, 5, 6, 6,$ and 5 respectively; and it is noted that ρ_u has less symmetry about zero, and the $\text{Min } (\rho_u)$ is around -0.2 in these cases. Comparison of Figures 16 and 20 indicates that gamma and truncated normal densities in which the corresponding means and variances of a_j and a_j' are equal give rise to very similar correlograms. The damping effect is slightly more marked for the truncated normal densities. Inspection of Figure 22 indicates that the symmetry property of the correlogram is not destroyed when one of the variables, a_j, a_j' , has zero variance so long as the means of a_j and a_j' are equal.

Equation (34) may be rewritten in terms of the symbols employed in this analysis. The result is given below:

$$\bar{p}_u = \frac{2}{n(n-1)} \sum_{\delta} (n-\delta) \rho_{\delta u} ,$$

and since from Equation (30),

$$\frac{\sigma_x^2}{n} = \frac{\sigma^2}{n} [1 + (n-1) \bar{p}_u] ,$$

systematic sampling may be expected to be more precise than simple random sampling if

$$\sum_{\delta} (n-\delta) \rho_{\delta u} < 0 . \quad (136)$$

Analysis of the correlograms of Figures 7 - 24 with dividers indicates that inequality (136) is satisfied where $u < 3/4 E(a_j + a_j')$ and $C < 3$; and this hypothesis is tested at a later point. It is obvious from the graphs that $\bar{\rho}_u > 0$ if $u = E(a_j + a_j')$, and this constitutes the danger about which repeated warnings have been stated in the literature. In the case of the gamma densities where $\lambda = 5.0$, $\lambda' = 4.0$, $r = 30$, $r' = 16$, and the mean cycle length is 10.0, systematic sampling with one random start on $[0, 10.0]$ yields an estimate of P_T that has roughly 1.5 times the variance of the estimate obtainable from simple random sampling at the same intensity. This estimate of relative precision was determined from extrapolation of Figure 12.

The rate of damping is an interesting feature of the correlograms, since cases exist in practice in which damping is sufficient so that if u is large relative to the mean cycle length, then $\rho_{\delta u} \cong 0$ for $\delta = 1, 2, 3, \dots$; and as a result, $\bar{\rho}_u \cong 0$.

In the case of the gamma densities for a_j and a_j' , the two factors affecting the rate of damping seem to be the coefficient of variation in the cycle length and C . As CV gets large for $C = 1.0$, the rate of damping increases, and as C increases, the rate of damping increases even though the coefficient of variation is small. Figures 8 - 15 were analyzed in an attempt to determine the relationship between the height of the correlogram at the end of s cycles and the parameters C and CV. Each correlogram required about two hours of electronic computer time and the expense involved resulted in the number of

runs for fixed C being too small to draw strong inference of the general case; however, several models were tested for their representativeness. The models are given below where y_s = the value of ρ_u where $u = s [E(a_j + a_j')]$.

$$\begin{aligned}
 1. \quad y_s &= e^{-[\alpha_1 + \alpha_2(C) + \alpha_3(CV)]s} \\
 2. \quad y_s &= e^{-[\alpha_1 + \alpha_2(C^\beta) + \alpha_3(CV)]s} \\
 3. \quad y_s &= [C^{\alpha_1} e]^{-[\alpha_2(CV)]s} \\
 4. \quad y_s &= \frac{\sin 2\pi s}{[\alpha_1 + \alpha_2(C^\beta) + \alpha_3(CV)]s}
 \end{aligned}$$

Particular ones of these models gave good fit for particular correlograms, but constants $\alpha_1, \alpha_2, \alpha_3, \beta$ giving good fit for all cases could not be obtained; however, the following approximation was found to be good for the ranges of values C and CV employed herein.

$$y_s = [1 - \alpha_1(C^\beta)(CV)]^{s-1} [\alpha_2 + \alpha_3(C^\beta)(CV)] \quad (137)$$

The constants $\alpha_1 = 2.0$, $\beta = 0.21$, $\alpha_2 = 0.90$, and $\alpha_3 = -2.6$ give good predicted results. The comparison of simulation vs. predicted results is given in Table 5 for the first three cycles. It is carefully noted that Equation (114) is not a "true" model since $[C^{0.21}(CV)]$ must be less than 0.35 in order for the results to be sensical. If $[C^{0.21}(CV)] \geq 0.35$, $y_s = 0.0$ for $s = 1, 2, \dots$ by definition.

Table 5. Actual vs. Predicted Values for y_s
for Gamma Densities

| CV | C | y_1 | | y_2 | | y_3 | |
|-------|------|--------|-----------|--------|-----------|--------|-----------|
| | | Actual | Predicted | Actual | Predicted | Actual | Predicted |
| 0.100 | 1.00 | 0.67 | 0.64 | 0.54 | 0.53 | 0.45 | 0.42 |
| 0.112 | 1.67 | 0.56 | 0.57 | 0.45 | 0.43 | 0.34 | 0.32 |
| 0.150 | 1.00 | 0.50 | 0.51 | 0.31 | 0.36 | 0.18 | 0.24 |
| 0.112 | 4.00 | 0.50 | 0.51 | 0.34 | 0.36 | 0.24 | 0.24 |
| 0.149 | 1.50 | 0.48 | 0.48 | 0.34 | 0.32 | 0.22 | 0.22 |
| 0.177 | 1.00 | 0.45 | 0.44 | 0.24 | 0.28 | 0.14 | 0.17 |
| 0.224 | 1.00 | 0.32 | 0.32 | 0.10 | 0.17 | 0.02 | 0.06 |
| 0.212 | 2.50 | 0.26 | 0.27 | 0.14 | 0.14 | 0.07 | 0.07 |

In the case of the truncated normal densities, Figures 16 - 24 were analyzed in much the same manner as the correlograms resulting from the gamma densities. The same four models as previously tested were again found inadequate, and Equation (137) with parameters $\alpha_1 = 2.0$, $\beta = 0.20$, $\alpha_2 = 0.8$, and $\alpha_3 = -2.1$ was found to be a better approximation than any model tested. Actual vs. predicted results are given in Table 6 for the first three cycles.

In practical situations the activity state of interest sometimes occurs as an element of a work cycle. If, for example, $E(a_j + a_j') = 1.0$ minute, $C = 1$, the densities are gamma, $CV = 0.1$, and an interval of 20 minutes is employed in a systematic sampling design, then $p_{20} \cong 0.0075$, $p_{40} \cong 0.00000056$, $p_{60} \cong 0.0$, ..., and

Table 6. Predicted vs. Actual Values for y_s
for Truncated Normal Distributions

| CV | C | y_1 | | y_2 | | y_3 | |
|-------|-----|--------|-----------|--------|-----------|--------|-----------|
| | | Actual | Predicted | Actual | Predicted | Actual | Predicted |
| 0.071 | 1.0 | 0.76 | 0.65 | 0.68 | 0.56 | 0.63 | 0.47 |
| 0.003 | 5.0 | 0.70 | 0.79 | 0.67 | 0.78 | ---- | 0.77 |
| 0.100 | 1.0 | 0.63 | 0.59 | 0.44 | 0.47 | 0.26 | 0.38 |
| 0.064 | 6.0 | 0.59 | 0.61 | ---- | 0.50 | ---- | 0.41 |
| 0.141 | 1.0 | 0.56 | 0.50 | 0.39 | 0.36 | 0.26 | 0.26 |
| 0.081 | 6.0 | 0.52 | 0.55 | ---- | 0.42 | ---- | 0.32 |
| 0.093 | 5.0 | 0.47 | 0.53 | 0.35 | 0.39 | ---- | 0.29 |
| 0.173 | 1.0 | 0.46 | 0.44 | 0.25 | 0.29 | 0.15 | 0.17 |
| 0.347 | 1.0 | 0.12 | 0.13 | 0.02 | 0.04 | 0.00 | 0.01 |

$$\frac{2}{n(n-1)} \sum_{\delta} (n-\delta) \rho_{\delta u} \approx 0.0, \quad (138)$$

If $C > 1$ in the above example, then $\rho_{\delta 20}$, $\delta = 1, 2, \dots$ is smaller than the indicated values. The solution of Equation (137) for s yields:

$$s = \left\{ \frac{\log y_s - \log [\alpha_2 + \alpha_3 (C^3)(CV)]}{\log [1 - \alpha_1 (C^3)(CV)]} \right\} + 1.0, \quad (139)$$

and the practitioner may employ this equation by specifying y_s , estimating C , CV , and $E(a_j + a_j')$, and selecting the interval d between observations so that

$$d > s E(a_j + a_j'). \quad (140)$$

The estimates referred to above are judgement estimates; however, if y_s is specified sufficiently small, error of estimation is unlikely to have serious effect, and $C = 1$ may be employed since this gives the worst condition from this standpoint.

Secondary Analysis. At this point, we consider a finite population of $nk = 960$, thirty second instants. It is noted that such populations were generated as particular realizations in the simulation study. Each set of parameters gave rise to fifty such populations. Each population of this type may be divided into all possible systematic samples for a particular k value as shown in Table 7. It is convenient to let y_{ij} represent the value of the j^{th} observation in the i^{th} possible systematic sample, so that $j = 1, 2, 3, \dots, n$, and $i = 1, 2, 3, \dots, k$. The population mean and variance are redefined in terms of y_{ij} as

$$p_N = \frac{1}{nk} \sum_{i=1}^k \sum_{j=1}^n y_{ij} \quad (141)$$

and

$$\sigma_N^2 = \frac{1}{nk} \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - p_N)^2 = p_N(1 - p_N) . \quad (142)$$

Employing the fundamental identity of the analysis of variance,

$$\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - p_N)^2 = n \sum_{i=1}^k (\bar{y}_{i.} - p_N)^2 + \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 , \quad (143)$$

where $\bar{y}_{i.}$ is the mean of the i^{th} possible systematic sample. It follows that

Table 7. Possible Systematic Samples

| Sample Number | | | | |
|-------------------|----------------|----------------------|----------------------|-------------|
| 1 | 2 | 3 ... | i ... | k |
| x_1 | x_2 | $x_3 \dots$ | $x_i \dots$ | x_k |
| x_{k+1} | x_{k+2} | $x_{k+3} \dots$ | $x_{k+i} \dots$ | x_{2k} |
| x_{2k+1} | x_{2k+2} | $x_{2k+3} \dots$ | $x_{2k+i} \dots$ | x_{3k} |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| $x_{(n-1)k+1}$ | $x_{(n-1)k+2}$ | $x_{(n-1)k+3} \dots$ | $x_{(n-1)k+i} \dots$ | x_{nk} |
| Means \bar{x}_1 | \bar{x}_2 | $\bar{x}_3 \dots$ | $\bar{x}_i \dots$ | \bar{x}_k |

$$\sigma_{\bar{y}_{i.}}^2 = \sigma_N^2 - \frac{\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2}{nk}, \quad (144)$$

since by definition

$$\sigma_{\bar{y}_{i.}}^2 = \frac{\sum_{i=1}^k (\bar{y}_{i.} - p_N)^2}{k}. \quad (145)$$

The sample mean obtained from simple random sampling with replacement has variance

$$\sigma_{\bar{y}}^2 = \frac{\sigma_N^2}{n} \quad (146)$$

and it follows that a necessary and sufficient condition for systematic sampling to be more precise than simple random sampling at the same sampling intensity

$$Q(k) = \frac{\sigma_{\bar{Y}_i.}}{\sigma_{\bar{Y}}} = \sqrt{n - \frac{\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2}{k \sigma_N^2}} < 1.0 . \quad (147)$$

This statement is relative to a particular finite population and a given value for k . Integers n and k must be selected so that $nk = N$, the number of instants in the population.

It was stated earlier that investigation of the correlograms seems to indicate that systematic sampling with $k < (3/4)E(a_j + a_j')$ gives an estimate of p_N having less variance than the estimate obtainable from simple random sampling at the same intensity when $C < 3$ even though the population structure is strongly periodic. This hypothesis was tested by selecting three sets of simulated activity, each consisting of 50 realizations, and calculating $Q(k)$ for k values of: 2,3,4,5,6,8,10,12,15, 16,20,24,30,32,40,48, and 60 instants. Two sets of these selected activity simulations resulted from gamma densities of a_j and a_j' , and one set of activity simulations resulted from truncated normal densities of the span lengths. The selected activity sets were for $C = 1.0$, $CV = 0.10$; $C = 2.33$, $CV = 0.212$; and $C = 1.0$, $CV = 0.173$. Two additional sets were selected in which $C > 3$. They were for $C = 4.0$, $CV = 0.177$; and $C = 5$, $CV = 0.003$. Average and maximum values of $Q(k)$ averaged over the fifty realizations are given in Tables 8-A and 8-B for the

Table 8-A. Average and Maximum Values of Q(k) for
Gamma Distributed Span Lengths

| k(30 sec. Instants) | $\lambda=\lambda'=5.0, r=r'=50.0$ Mean Cycle Length 20 Min. (40 instants) C=1.0, CV = 0.10 | | | $\lambda=3.0, \lambda'=2.0, r=9, r'=14$ Mean Cycle Length 10 Min. (20 instants) C=2.33, CV=0.212 | | | $\lambda=\lambda'=4.0, 4=64, r'=16$ Mean Cycle Length 20 Min. (40 instants) C=4.0, CV=0.177 | | |
|------------------------|---|-------------|-----------------|---|-------------|-----------------|--|-------------|-----------------|
| | <u>Ave.</u> | <u>Max.</u> | <u>Std.Dev.</u> | <u>Ave.</u> | <u>Max.</u> | <u>Std.Dev.</u> | <u>Ave.</u> | <u>Max.</u> | <u>Std.Dev.</u> |
| | | | | | | | | | |
| 2 | 0.1373 | 0.4112 | 0.0973 | 0.1893 | 0.6012 | 0.1420 | 0.1575 | 0.4618 | 0.1250 |
| 3 | 0.1799 | 0.4184 | 0.0881 | 0.3177 | 0.8622 | 0.1494 | 0.2414 | 0.5895 | 0.1346 |
| 4 | 0.2291 | 0.6090 | 0.1089 | 0.3372 | 0.6374 | 0.1390 | 0.2939 | 0.6482 | 0.1343 |
| 5 | 0.2486 | 0.5813 | 0.1163 | 0.3696 | 0.6876 | 0.1387 | 0.3619 | 0.7205 | 0.1317 |
| 6 | 0.3026 | 0.6107 | 0.1238 | 0.4449 | 0.8926 | 0.1347 | 0.3653 | 0.5996 | 0.1088 |
| 8 | 0.3546 | 0.7008 | 0.1356 | 0.5025 | 1.2056 | 0.2046 | 0.3854 | 0.6785 | 0.1272 |
| 10 | 0.3370 | 0.5953 | 0.1147 | 0.6786 | 1.6047 | 0.2559 | 0.4557 | 0.7712 | 0.1305 |
| 12 | 0.4432 | 0.6749 | 0.1193 | 0.6527 | 1.1598 | 0.1976 | 0.6022 | 1.1940 | 0.1994 |
| 15 | 0.4312 | 0.6365 | 0.1138 | 0.7298 | 1.4379 | 0.2561 | 0.5730 | 1.0655 | 0.1691 |
| 16 | 0.4007 | 0.7293 | 0.1126 | 0.8373 | 1.7309 | 0.2949 | 0.6526 | 1.2261 | 0.2217 |
| 20 | 0.3785 | 0.5949 | 0.1014 | 1.2000 | 2.3473 | 0.4538 | 0.9811 | 2.0666 | 0.4559 |
| 24 | 0.4571 | 0.7344 | 0.0978 | 0.8931 | 1.6791 | 0.2383 | 0.6738 | 1.1132 | 0.1647 |
| 30 | 0.5249 | 1.0604 | 0.1459 | 0.8106 | 1.2428 | 0.1759 | 0.6361 | 1.1023 | 0.1446 |
| 32 | 0.5831 | 1.2211 | 0.1918 | 0.8162 | 1.4242 | 0.1774 | 0.7040 | 0.9742 | 0.1439 |
| 40 | 2.4815 | 4.1472 | 0.9588 | 1.0197 | 1.7837 | 0.2803 | 1.8264 | 2.9956 | 0.6086 |
| 48 | 0.6710 | 1.0838 | 0.1955 | 0.8733 | 1.4167 | 0.1730 | 0.8006 | 1.3091 | 0.1572 |
| 60 | 0.5587 | 0.9083 | 0.1058 | 0.9682 | 1.4907 | 0.2164 | 0.8357 | 1.4633 | 0.1951 |

Table 8-B. Average and Maximum Values of
Q(k) for Span Lengths with
Truncated Normal Distributions

| k(30 sec. Instants) | $\mu=\mu'=10.0, \sigma^2=\sigma'^2=6.0$ Approx. Mean Cycle Length 20 Min. (40 instants) $C=1.0, CV \approx 0.173$ | | | $\mu=4.0, \mu'=20.0, \sigma^2=\sigma'^2=0.25$ Approx. Mean Cycle Length 24 Min. (48 instants) $C=5.0, CV \approx 0.003$ | | |
|------------------------|--|-------------|-----------------|--|-------------|-----------------|
| | <u>Ave.</u> | <u>Max.</u> | <u>Std.Dev.</u> | <u>Ave.</u> | <u>Max.</u> | <u>Std.Dev.</u> |
| 2 | 0.1412 | 0.3200 | 0.0973 | 0.1702 | 0.4975 | 0.1087 |
| 3 | 0.1860 | 0.4566 | 0.0993 | 0.2456 | 0.7468 | 0.1346 |
| 4 | 0.2311 | 0.3811 | 0.0769 | 0.2621 | 0.5962 | 0.1233 |
| 5 | 0.2458 | 0.5324 | 0.0977 | 0.2798 | 0.5449 | 0.1128 |
| 6 | 0.3042 | 0.5460 | 0.1262 | 0.5172 | 0.9583 | 0.1863 |
| 8 | 0.3625 | 0.6387 | 0.1133 | 0.2380 | 0.4988 | 0.0938 |
| 10 | 0.3735 | 0.6895 | 0.1109 | 0.3351 | 0.4868 | 0.0848 |
| 12 | 0.4359 | 0.9546 | 0.1380 | 1.2564 | 2.1456 | 0.5042 |
| 15 | 0.4788 | 0.9092 | 0.1705 | 0.4433 | 0.8205 | 0.1275 |
| 16 | 0.4911 | 0.8304 | 0.1411 | 1.9514 | 2.9929 | 0.7099 |
| 20 | 0.4978 | 0.8085 | 0.1508 | 0.3553 | 0.4886 | 0.0629 |
| 24 | 0.5135 | 0.7598 | 0.1107 | 2.7344 | 3.5332 | 0.5766 |
| 30 | 0.5677 | 0.9974 | 0.1605 | 0.4614 | 0.7627 | 0.0818 |
| 32 | 0.6989 | 1.3055 | 0.2738 | 1.4117 | 2.0814 | 0.4858 |
| 40 | 1.5001 | 3.0173 | 0.4416 | 0.3652 | 0.6998 | 0.0761 |
| 48 | 0.9207 | 2.1649 | 0.3279 | 3.3870 | 4.0526 | 0.4777 |
| 60 | 0.6531 | 0.9464 | 0.1300 | 0.7097 | 1.0073 | 0.1624 |

gamma and truncated normal span lengths respectively. The standard deviations are also given. The average and standard deviation were calculated as follows:

$$\overline{Q(k)} = \frac{1}{50} \sum_{i=1}^{50} [Q(k)]_i, \quad (148)$$

and

$$\text{Std. Dev. of } Q(k) = \sqrt{\frac{\sum_{i=1}^{50} \{[Q(k)]_i - \overline{Q(k)}\}^2}{49}}. \quad (149)$$

Histograms of the $Q(k)$ values are presented in Appendix B, Figures 37-41 for the gamma and truncated normal densities of span length.

In the three cases where $C < 3$, it is noted that the average value of $Q(k)$ is less than one when $k \leq (3/4)E(a_j + a_j')$; therefore we conclude that on the average systematic sampling is more precise than random sampling if $C < 3$, $k \leq (3/4)E(a_j + a_j')$, and the coefficient of variation in cycle length is roughly $1/10$ to $1/5$. It is noted in the cases where $C = 1$ that the maximum $Q(k)$ is less than 1.00 in all cases where $k < (3/4)E(a_j + a_j')$. In the case where $C = 2.33$, the maximum values for $Q(k)$ exceeded one when $k \geq (2/5)E(a_j + a_j')$; however the histograms indicate no "bad behavior" for any of these values, k .

In the case where $C = 4.0$, it is noted that the average values of $Q(k)$ were less than one when $k \leq (3/4)E(a_j + a_j')$; however this case was selected due to the fact that the coefficient of variation is as large as 0.177.

The case where $C = 5.0$ and $CV = 0.003$ represents a very strong periodic situation, and although systematic sampling is superior for some values of k , the performance is erratic and the histograms displayed in Figure 41 show less central tendency.

The unusual class intervals employed for the histograms of Figures 37-41 resulted from the use of electronic computing equipment and pre-established rules for the determination of the intervals. The computer programs for the calculation of $Q(k)$, related values, and the histograms are given in Appendix B, Table 16.

Conclusions

When the span lengths have negative exponential densities, systematic sampling with one random start yields an estimate having less variance than the estimate obtainable from simple or stratified random sampling at the same intensity.

In the cases where the span lengths have gamma or truncated normal densities, the ratio of the standard deviation of the systematic sampling estimate to that of simple random sampling for particular values of C and CV is dependent upon the relationship of the interval between observations to the mean cycle length. If the interval between observations is large relative to the mean cycle length, Equations (139) and (140) lead to an indication of the safe zone in which the chance selection of an interval which is an integral multiple of the mean cycle length has only a small effect on the relative precision of the estimates.

If the interval between observations is less than $(3/4)$ of $E(a_j + a'_j)$ and $C < 3$, the systematic design seems to be superior for coefficient of variation values greater than $1/10$; and in all cases, including extreme cases, systematic sampling seems to outperform random sampling when the interval between observations is less than $(1/4) E(a_j + a'_j)$.

Problems associated with the estimation of the variance of the estimate are treated in the following chapter.

CHAPTER VIII

ESTIMATION WITH MULTIPLE OCCURRENCE OF THE ACTIVITY

Introduction

The objective of this chapter is to treat the problem of estimation of the variance of the sample mean. In the presentation of Chapter VII, it was noted that one difficulty associated with systematic sampling is the estimation of the variance of the estimate of p , p_T , or p_N . In the case where the process is Markovian, the consolation that systematic sampling yields an estimate having less variance than the estimate obtainable from random sampling has some value; however, in periodic situations in which the span lengths, a_j , a_j' have gamma, normal, or some other distributions, the relationship of the variances of the estimates from systematic and simple random sampling is quite dependent upon the interval between observations.

One Random Start

An Upper Bound to the Variance of the Estimate

With Knowledge of the Process. In the case where the process is Markovian, an overestimate of the variance of $\hat{p}_{sy.d}$ is given by

$\hat{\sigma}_{\hat{p}_{sy.d}}^2$ where

$$\hat{\sigma}_{\hat{p}_{sy.d}}^2 = \frac{\hat{p}_{sy.d} (1 - \hat{p}_{sy.d})}{n} . \quad (150)$$

Since $E(\hat{p}_{sy.d}) = p$ due to the random start, then

$$E(\hat{\sigma}_{\hat{p}_{sy.d}}^2) = \left[\frac{p(1-p)}{n} \right] - \frac{1}{n} \left[\sigma_{\hat{p}_{sy.d}}^2 \right], \quad (151)$$

and if we let $K = \sigma_{\hat{p}_{ran}}^2 / \sigma_{\hat{p}_{sy.d}}^2$, then

$$E(\hat{\sigma}_{\hat{p}_{sy.d}}^2) = \left(K - \frac{1}{n} \right) \sigma_{\hat{p}_{sy.d}}^2. \quad (152)$$

As K increases this estimate becomes more positively biased.

In the case where the process is periodic, there is a gamble associated with the use of Equation (150) for the estimation of variance. If $d = E(a_j + a_j')$, the variance will be considerably greater than estimated. If, however, there is knowledge of the factors $E(a_j + a_j')$ and C , then Equation (150) can be expected to give conservative results when $d < (3/4) E(a_j + a_j')$ and $C < 3$. If $C > 3$, the value of $d = (1/2) E(a_j + a_j')$ should be avoided as an interval between observation. In the case where s as defined in Equation (139) is large enough so that Inequality (140) is satisfied, then the practitioner is again justified in the use of Equation (150).

The practitioner may wish to continuously time several activity cycles in order to estimate $E(a_j)$ and $E(a_j')$. This study procedure was advocated by Flowerdew and Malin (58, p. 206). It is noted that the use of Equation (150) does not permit the practitioner to state the advantage in the reduction in variance due to the use of a systematic design.

With No Knowledge of the Process. If the sampling is to extend over twenty periods or more where a shift or half shift constitutes a period, and no assumptions about the population structure are permitted, we shall consider the population as stratified by period with each period forming a stratum. A single realization of the process is considered, so that $p_{T,h}$ is the fraction of time spent in activity A during period h, where $h = 1, 2, 3, \dots, L$. Systematic sampling with a single independent random start in each stratum is employed, and the estimate of $p_{T,h}$ is $p_{sy.d.h}$, where d is the interval between observations in period h. The value $d = \text{constant}$, for $h = 1, 2, \dots, L$. The fraction of the entire period, LT, of study that is covered by activity A is p_{LT} , and this quantity was defined earlier in Equation (121). The estimate of p_{LT} is denoted by \hat{p}_{LT} , where

$$\hat{p}_{LT} = \frac{1}{L} \sum_{h=1}^L \hat{p}_{sy.d.h}, \quad (153)$$

and this estimate is obviously unbiased since the estimates within each stratum are unbiased.

Cochran (39, p.183) has considered an estimate of this form, so that we may extend his result to attribute type measurement in order to state an estimate of the variance of \hat{p}_{LT} that is on the average an over estimate. The resulting estimate is $\hat{\sigma}_{\hat{p}_{LT}}^2$, where

$$\hat{\sigma}_{\hat{p}_{LT}}^2 = \frac{1}{L^2} \sum_{\text{pairs}(h,j)} (\hat{p}_{sy.d.h} - \hat{p}_{sy.d.j})^2. \quad (154)$$

where there are $L/2$
pairs with distinct
membership

The L strata are divided into $L/2$ pairs, established before the sampling is conducted, so that stratum i is a member of only one pair. For example, if four strata are denoted as st_1 , st_2 , st_3 , and st_4 , possible sets of pairs are $\{(st_1, st_2), (st_3, st_4)\}$, $\{(st_1, st_3), (st_2, st_4)\}$, $\{(st_1, st_4), (st_2, st_3)\}$. One of the three sets is selected, based on the criterion suggested by Equation (155), as a domain of summation.

Since

$$E(\hat{\sigma}_{\hat{p}_{LT}}^2) = V(\hat{p}_{LT}) + \frac{1}{L^2} \sum_{\text{pairs}(h,j)} (p_{T,h} - p_{T,j})^2 \quad (155)$$

it is advisable to select strata for a particular pair in which $p_{T,h}$ and $p_{T,j}$ are nearly equal.

Cochran states:

Where stratification with numerous strata is employed, and an independent systematic sample is drawn from each stratum, the effects of any hidden periodicities tend to cancel out, and an estimate of error which is known to be an overestimate can be obtained (39, p. 185).

It is interesting to note that there are $(L-1)(L)/4$ possible sets of pairs of which the practitioner must select one set in order to estimate the variance. Knowledge of the strata means will be helpful in this selection, and once the selection is made, calculations are permitted. If there is a trend in the values $p_{T,h}$, it is logical to form pairs in order of occurrence in order to minimize the bias in the estimate of variance. If there is a cyclic tendency within weeks, it is logical to pair Mondays, Tuesdays, Wednesdays, etc.

In the limiting situation where strata are made successively smaller until there is only one observation per stratum, the sampling design reduces to a stratified random design with one observation per stratum.

Studies conducted by industrial engineers frequently extend over several months so that L may be quite large for either shift or half-shift strata. Utilizing the result stated in Equation (154), an approximate confidence interval on p_{LT} is given by

$$p \{ \hat{p}_{LT} - K_{\alpha/2} \hat{\sigma}_{\hat{p}_{LT}} \leq p_{LT} \leq \hat{p}_{LT} + K_{\alpha/2} \hat{\sigma}_{\hat{p}_{LT}} \} \approx 1 - \alpha, \quad (156)$$

where the estimates are assumed to be approximately normally distributed.

Multiple Random Starts

Study Over One Period

In order to obtain an unbiased estimate of the variance of the estimate of the population characteristic, p_T , associated with one time period, multiple random starts must be employed if a systematic sampling design is utilized. The sample mean associated with the i^{th} random start is denoted by $p_{\text{sy.d.i}}$, where d is the interval width between successive observations in a particular one of the g systematic samples, and

$$\hat{p}_{\text{sy.d.i}} = \frac{1}{n} \sum_{j=1}^n x[t_1 + (j-1)d]. \quad (157)$$

The value t_1 is the instant on $[0, d]$ randomly selected for the initial observation, and $nd = T$, the length of the period. The estimate of

p_T is $\hat{p}_{sy.d}$, where

$$\hat{p}_{sy.d} = \frac{1}{g} \sum_{i=1}^g \hat{p}_{sy.d.i} \quad (158)$$

which is obviously unbiased. An unbiased estimate of the variance of

$\hat{p}_{sy.d}$ is $\frac{\hat{\sigma}_{\hat{p}_{sy.d}}^2}{\hat{p}_{sy.d}}$, where

$$\hat{\sigma}_{\hat{p}_{sy.d}}^2 = \frac{1}{g(g-1)} \sum_{i=1}^g (\hat{p}_{sy.d.i} - \hat{p}_{sy.d})^2. \quad (159)$$

This result follows from the fact that an unbiased estimate of the variance of $\hat{p}_{sy.d.i}$ is given by

$$\hat{\sigma}_{\hat{p}_{sy.d.i}}^2 = \frac{1}{g-1} \sum_{i=1}^g (\hat{p}_{sy.d.i} - \hat{p}_{sy.d})^2. \quad (160)$$

An approximate confidence interval may be constructed on the parameter p_T as follows; where $\hat{p}_{sy.d.i}$ is assumed to have a normal distribution, so that $t_{\alpha/2, g-1}$ can be read from a table of the Student t distribution, with $g-1$ degrees of freedom.

$$p \{ \hat{p}_{sy.d} - t_{\alpha/2, g-1} (\hat{\sigma}_{\hat{p}_{sy.d}}) \leq p_T \leq \hat{p}_{sy.d} + t_{\alpha/2, g-1} (\hat{\sigma}_{\hat{p}_{sy.d}}) \} \quad (161)$$

$$\approx 1 - \alpha.$$

It is noted that ng observations are made for the period of length T .

Study Over L Periods

If L periods such as shifts or half-shifts form strata, and

ng observations are made within each stratum, each being of equal size, then $L.n.g$ observations are made for the entire study, and the parameter to be estimated is p_{LT} as defined in Equation (92). Strata of unequal sizes are not considered. An unbiased estimate of p_{LT} is \hat{p}_{LT} , where

$$\hat{p}_{LT} = \frac{1}{L} \sum_{h=1}^L \hat{p}_{sy.d.h} , \quad (162)$$

and $\hat{p}_{sy.d.h}$ is the estimate obtained in the h^{th} period or stratum.

An unbiased estimate of the variance of \hat{p}_{LT} is $\hat{\sigma}_{p_{LT}}^2$, where

$$\hat{\sigma}_{p_{LT}}^2 = \frac{1}{L^2} \frac{1}{g(g-1)} \sum_{h=1}^L \sum_{i=1}^g (\hat{p}_{sy.d.h.i} - \hat{p}_{sy.d.h})^2 \quad (163)$$

(39, p. 72) .

In this case, an approximate confidence interval on p_{LT} may be constructed as follows:

$$P \{ \hat{p}_{LT} - t_{\alpha/2, g-1} (\hat{\sigma}_{\hat{p}_{LT}}) \leq p_{LT} \leq \hat{p}_{LT} + t_{\alpha/2, g-1} (\hat{\sigma}_{\hat{p}_{LT}}) \} \quad (164)$$

$$\cong 1 - \alpha .$$

Concluding Remarks

There are several reasons why the practicing industrial engineer will find systematic sampling with multiple random starts to be a practical and useful design. In large studies where several observers are employed, and each "makes a round" by walking through the area in which the activity is being conducted, the interval between observation

of a particular subject by a particular observer is d ; and this time may be established so as to conveniently allow the observer to make his round. If each observer selects an instant at random on $[0, d]$ at which to make his first observation, and all observers observe the same subjects, then the results may be pooled as specified in Equation (158). Equation (159) may be employed for the estimation of variance. The estimate of variance becomes more precise as the number of observers increases. This chapter deals only with inference about single subjects; thus the above procedure does not imply that observational data should be pooled for all subjects.

In situations where only one observer is employed in conducting the study, multiple random starts will be less convenient than a single random start in so far as planning observation times is concerned; however, the regularity of spacing between observations for each interval of width d , will allow the observer to develop a pattern of observation and alternate work if such work exists.

From an intuitive standpoint, systematic sampling with multiple random starts has an appeal similar to that of stratified random sampling in that both designs distribute the observations more evenly across the population. Another advantage over systematic sampling with one random start is that human subjects are less likely to anticipate observation times.

CHAPTER IX

STUDY OF NATURAL POPULATIONS

Introduction

Several industries and service organizations were asked to participate in this study by providing data from continuous production study to be utilized to establish populations on which to test systematic sampling designs. It was agreed that the identity of the organization would not be revealed; however, only two organizations provided data. Fortunately, each of the participating firms provided several studies, so that different kinds of activity are represented.

The study code is a two digit integer the first of which identifies the firm, and the second identifies the study. The studies extended over 480 minutes each, and activity A, the state of interest is idle time. The study codes are 11, 12, 13, 14, 15, 16, and 21, 22, 23, and 24. The subjects of studies 11 and 12 were "floaters" working in steel fabrication. These floaters engage in an irregular type of work, assisting lay-out men. The demand for their service is irregular, and the duration of work and idle spans is irregular from occurrence to occurrence. Study 13 is a study of the activity of a "lay-out man" who lays out and marks steel components for welding. The lay-out man works in conjunction with a welder. Study 14 is a study of the activity of an overhead crane operator who operated a crane to move heavy, steel components and assemblies. Studies 15 and 16 are studies of another overhead crane operator on two consecutive days. Studies 21 and 22 were studies

of the activity of a mill and warehouse timekeeper whose duties include basic timekeeping functions such as preparing punch cards, preparing reports, checking clock cards to verify attendance, and collecting and distributing clock cards. The studies were over two not consecutive days. Studies 23 and 24 were of the activity of a mill weigher whose duties consist of preparing punched cards, preparing production reports, weighing and recording production, and preparing tags for identification of production.

Presentation of Data

The data are presented in tabular form in Tables 42-49 of Appendix B, and summarized data are presented in histogram form in Figure 42 of Appendix B. The Pearson chi square "goodness of fit" test was employed where the theoretical density was taken as the exponential density:

$$\begin{aligned} f(t) &= \hat{\lambda} e^{-\hat{\lambda}t} ; \quad t \geq 0 \\ &= 0 \quad ; \quad t < 0 , \end{aligned} \quad (165)$$

where $\hat{\lambda}$ is an estimate of the mean rate of activity state change. The within cycle correlation between a_j and a_j' was calculated by the relation:

$$\text{Correlation} = \frac{\sum_{j=1}^K (a_j - \bar{a})(a_j' - \bar{a}')}{\sqrt{\sum_{j=1}^K (a_j - \bar{a})^2 \sum_{j=1}^K (a_j' - \bar{a}')^2}} . \quad (166)$$

where there are K cycles during the 480 minute period of activity; and

the mean cycle length was calculated by:

$$\text{mean cycle length} = 480/K . \quad (167)$$

The results of the tests and calculations described in the preceding paragraph are presented in Table 9.

Analysis

In order to analyze the natural population data presented in Tables 42 - 51 of Appendix B, each realization was observed at fifteen second intervals; so that a finite population of 1920, fifteen second instants was formed for each 480 minute period of study. Equations (75) and (147) were employed to analyze each study and the results are presented in Figures 25 - 34. The ratio $Q(k)$ was calculated for $k = 2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 20, 24, 30, 32, 40, 48, 64, 80, 96, 120$, instants. It is noted that the correlogram is calculated from only one realization in each case.

Inspection of the ratios $Q(k)$ indicates that for the values of k employed, systematic sampling is a good design for all of the natural population studies. In study 12, two values of $Q(k)$ were slightly greater than one, and in study 14, one value of $Q(k)$ is slightly greater than one; however no value of $Q(k)$ exceeded 1.02. The correlograms show a general downward tendency, and periodic behavior such as displayed in Figure 29 is weak.

The electronic computer program for analyzing the natural populations making the calculations presented in Table 10 and Figures 25 - 34 is given in Appendix B as Table 39.

Table 9. Summary of Calculations from Natural Populations

| Study Number | Number of Cycles K | Mean Cycle Length (Min.) | P_{480} | Correlation Between a_j, a_j' for $j = 1, 2, \dots$ | Test of Hypothesis of No. Sig. Dep. from Exp. Dist. a_j | Test of Hypothesis of No. Sig. Dep. from Exp. Dist. a_j' | Figure No |
|--------------|--------------------|--------------------------|-----------|---|---|--|-----------|
| 11 | 74 | 6.4000 | 0.6745 | 0.1666 | Reject | Reject | 30, 31 |
| 12 | 81 | 5.9259 | 0.5984 | -0.0449 | Reject | Reject | 30, 32 |
| 13 | 38 | 12.6316 | 0.5838 | -0.0352 | Accept | Accept | 30, 33 |
| 14 | 47 | 10.2128 | 0.5427 | -.1199 | Accept | Reject | 30, 34 |
| 15 | 36 | 13.3333 | 0.3260 | 0.2646 | Accept | Accept | 30, 35 |
| 16 | 30 | 16.0000 | 0.5484 | -0.0692 | Accept | Accept | 30, 36 |
| 21 | 16 | 30.0000 | 0.4781 | 0.0627 | Insufficient Data | Insufficient Data | 30, 37 |
| 22 | 13 | 36.9231 | 0.5331 | 0.3608 | Insufficient Data | Insufficient Data | 30, 38 |
| 23 | 55 | 8.7273 | 0.7172 | 0.1658 | Accept | Reject | 30, 39 |
| 24 | 46 | 10.4348 | 0.7760 | 0.0430 | Reject | Reject | 30, 40 |

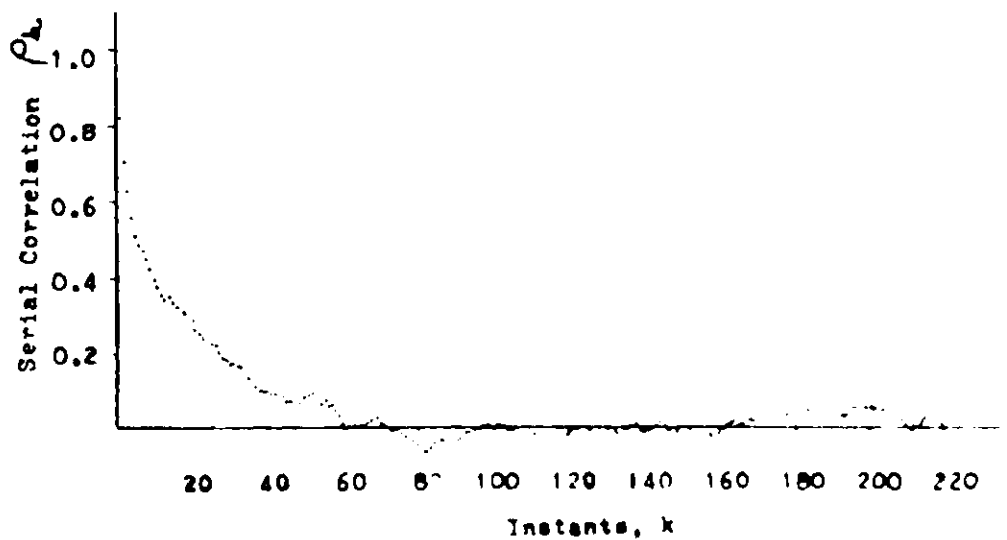
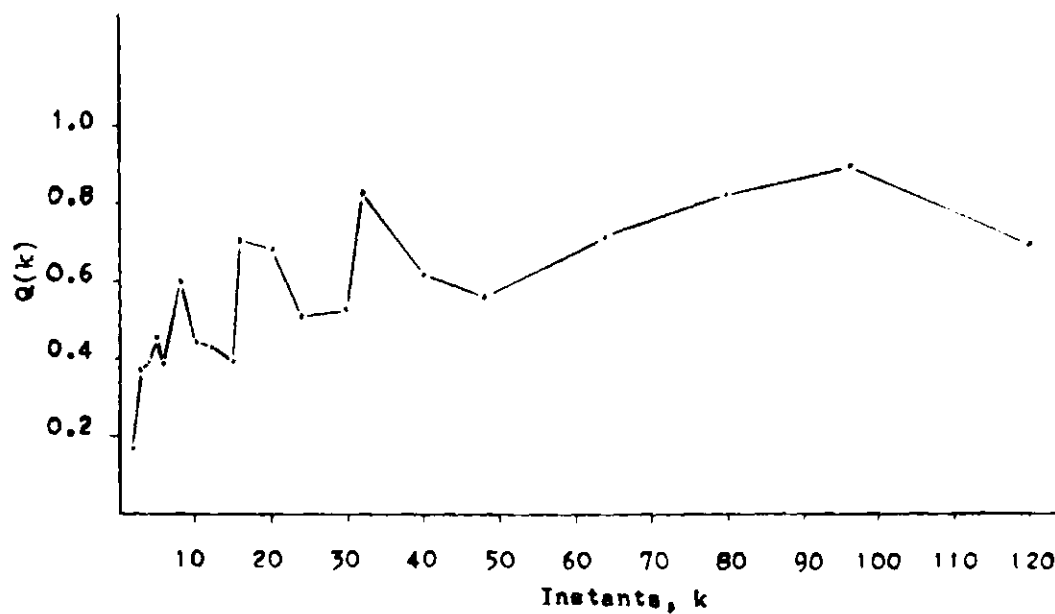


Figure 25. The Ratios $Q(k)$ and the Function ρ_k for Study No. 11

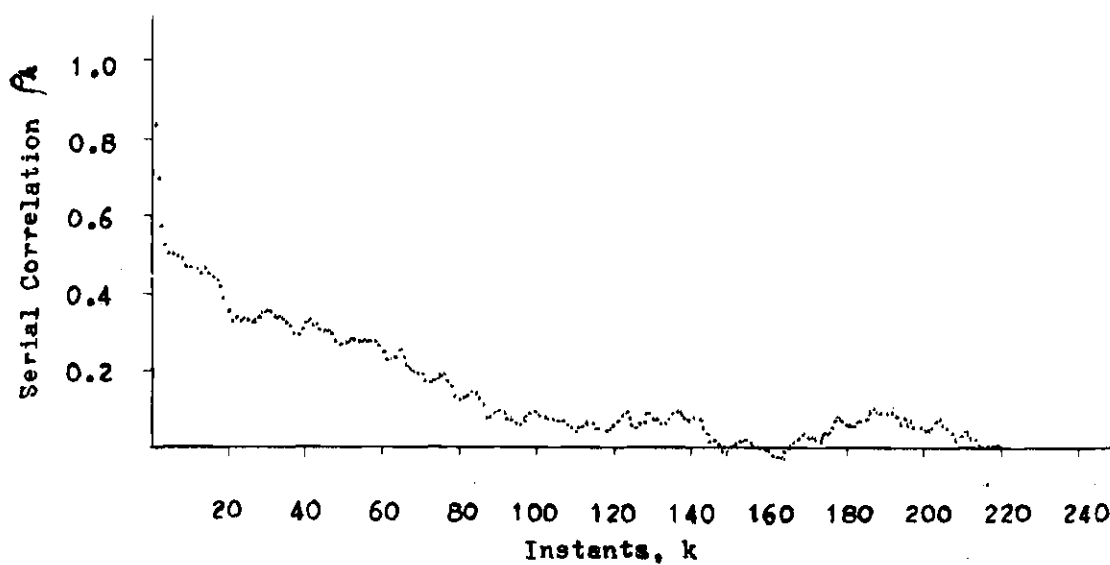
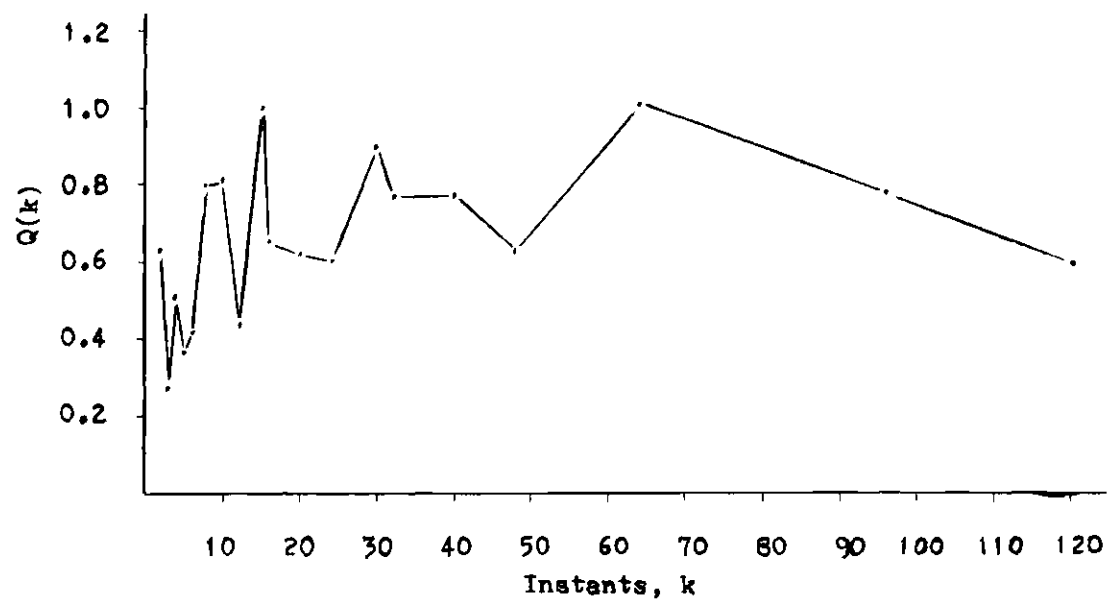


Figure 26. The Ratios $Q(k)$ and the Function ρ_k for Study No. 12

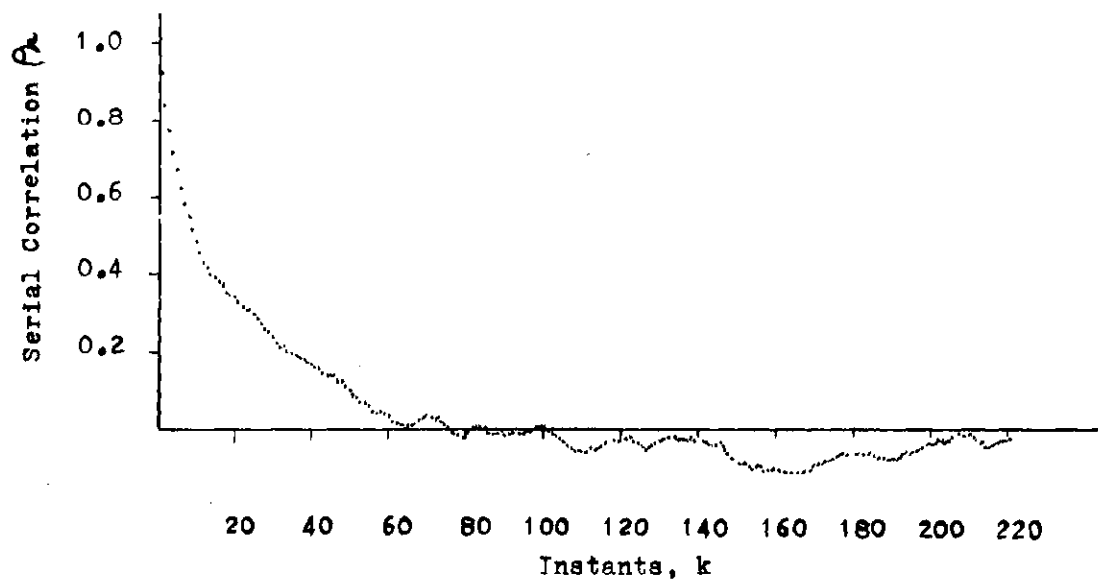
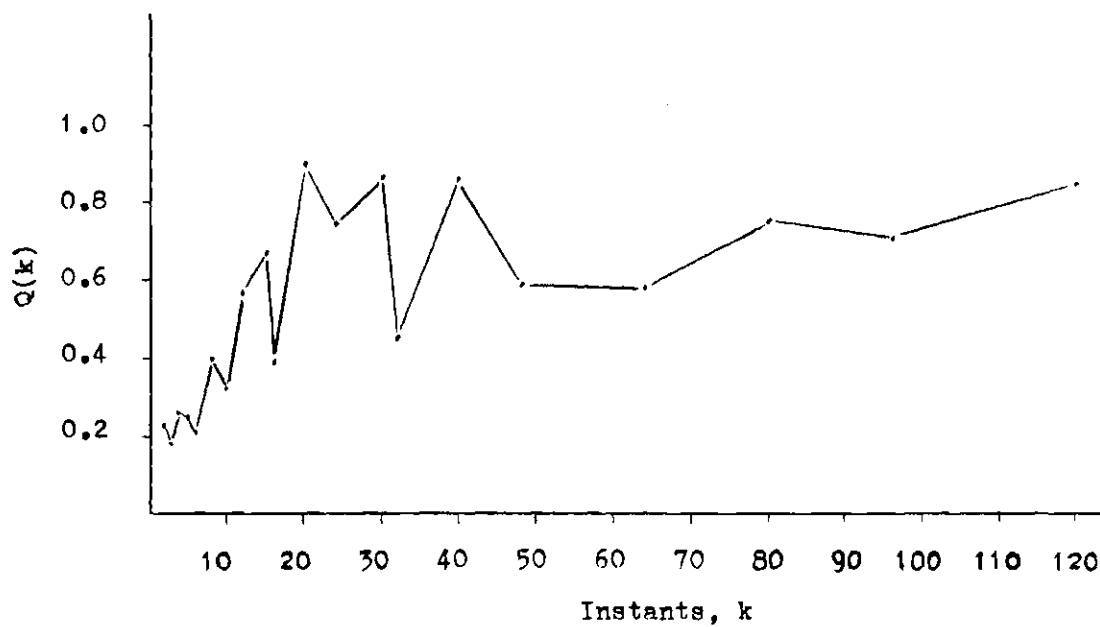


Figure 27. The Ratios $Q(k)$ and the Function ρ_k for Study No. 13

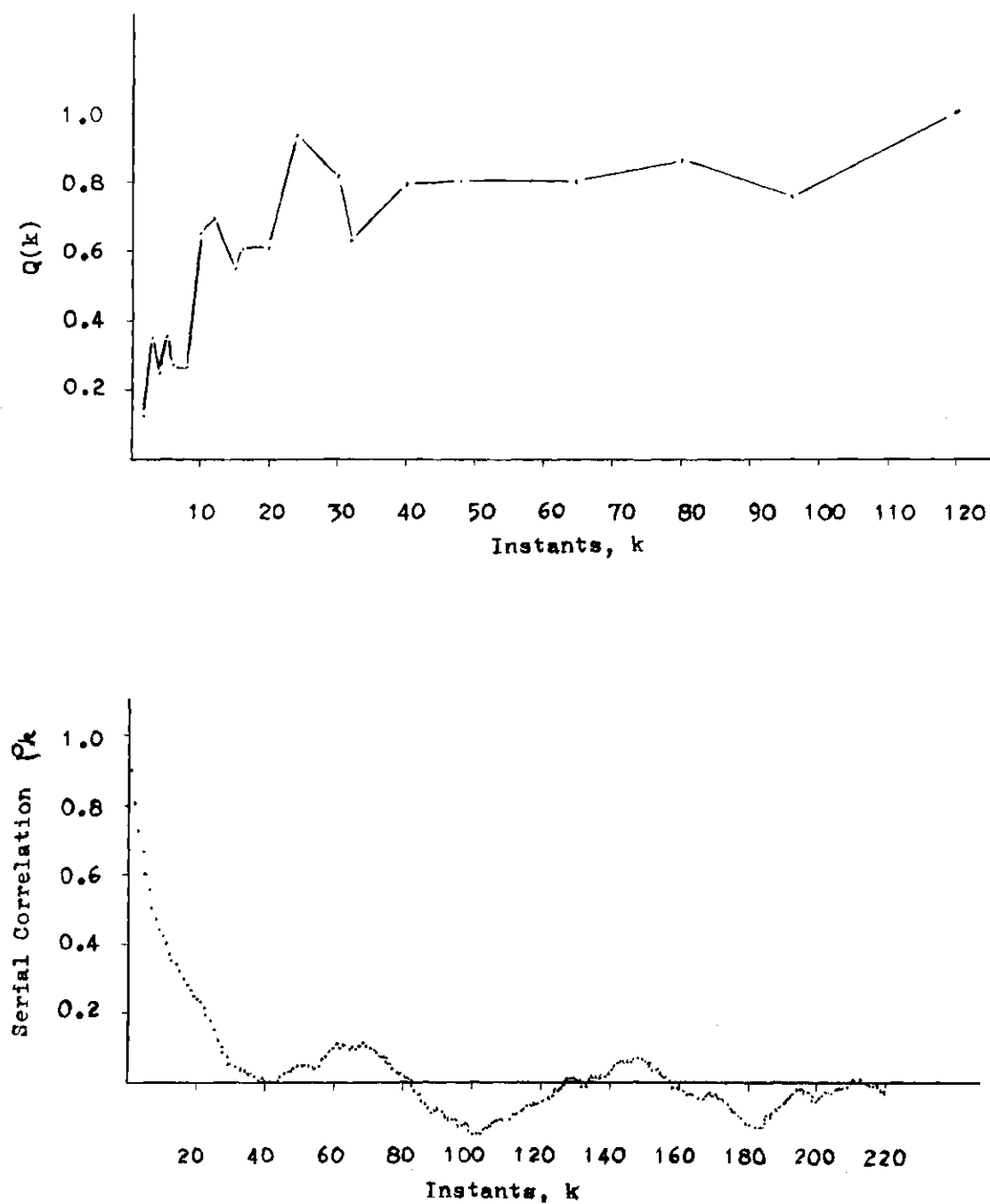


Figure 28. The Ratios $Q(k)$ and the Function ρ_k for Study No. 14

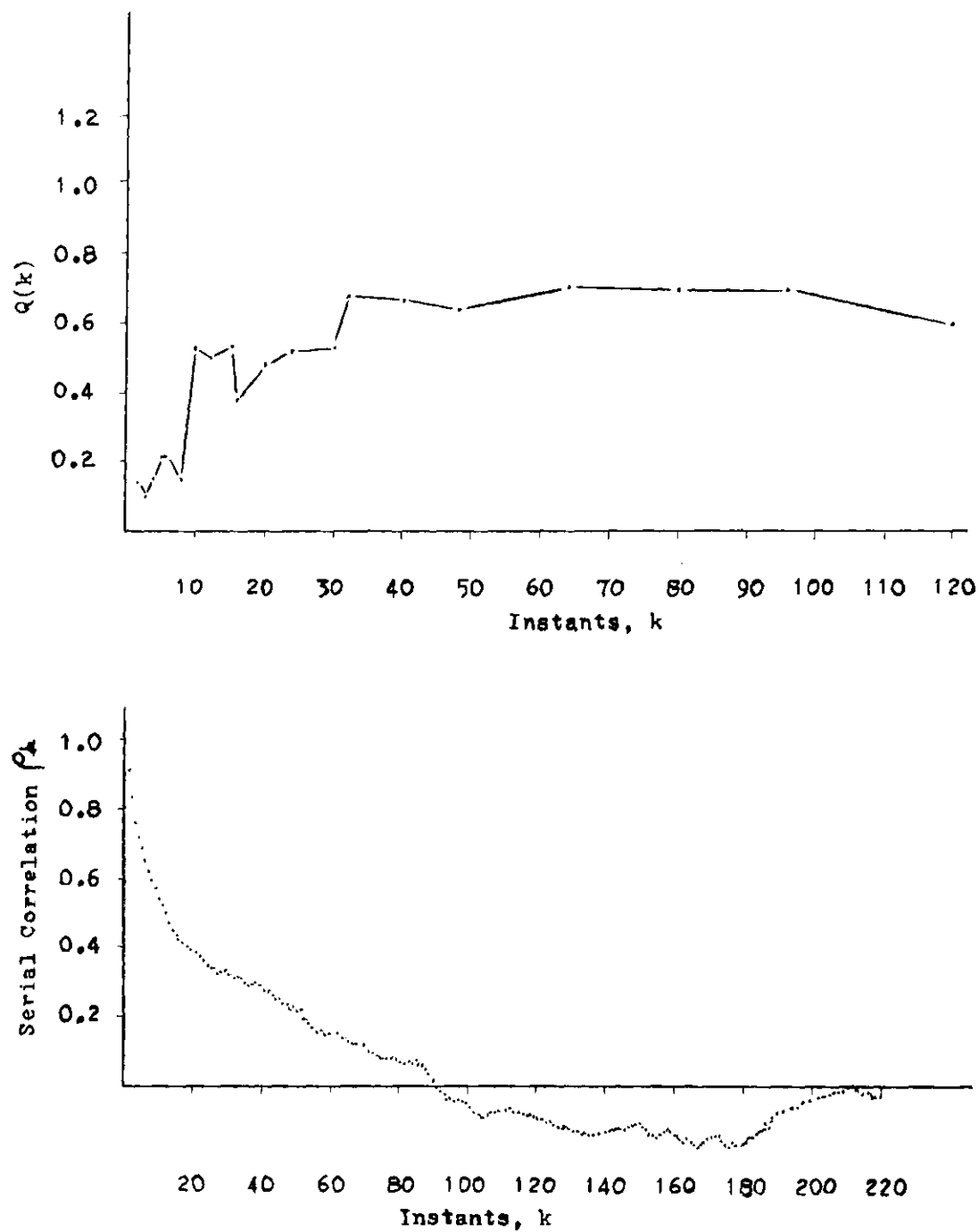


Figure 29. The Ratios $Q(k)$ and the Function ρ_k for Study No. 15

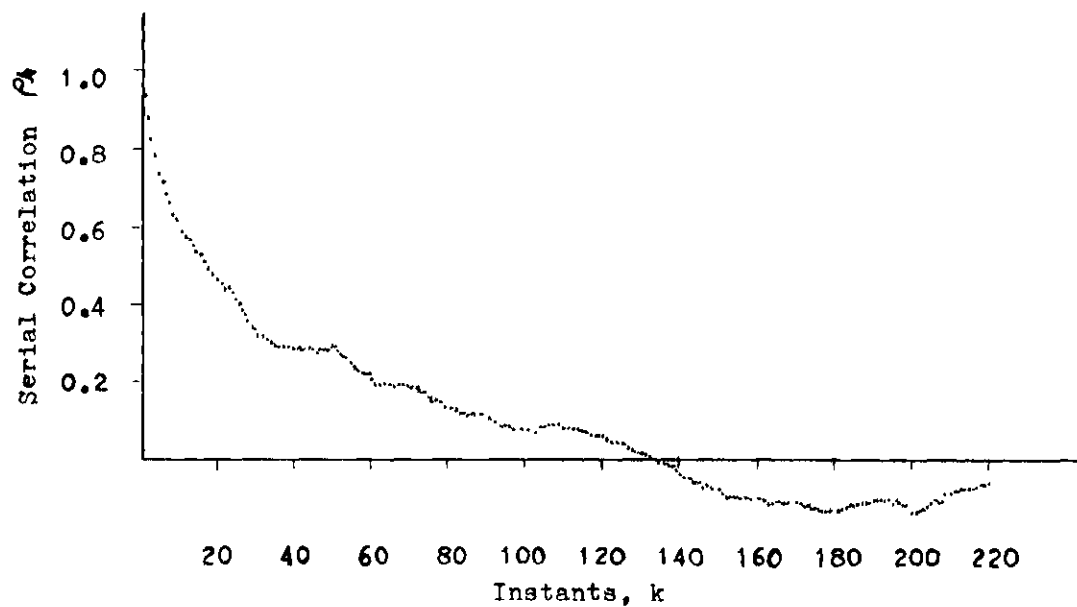
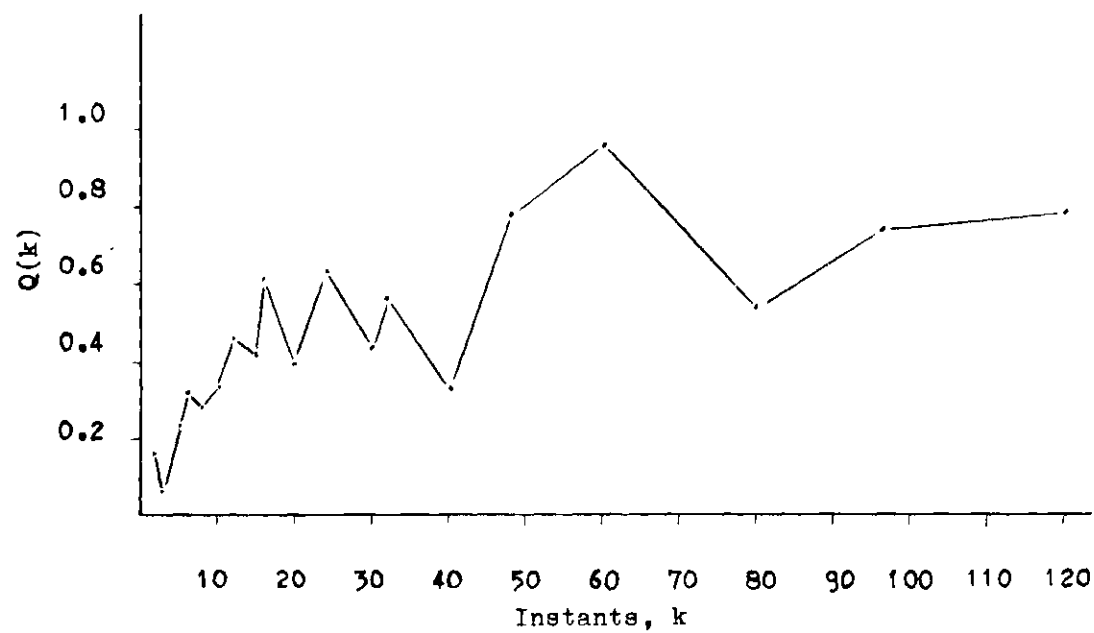


Figure 30. The Ratios $Q(k)$ and the Function ρ_k for Study No. 16

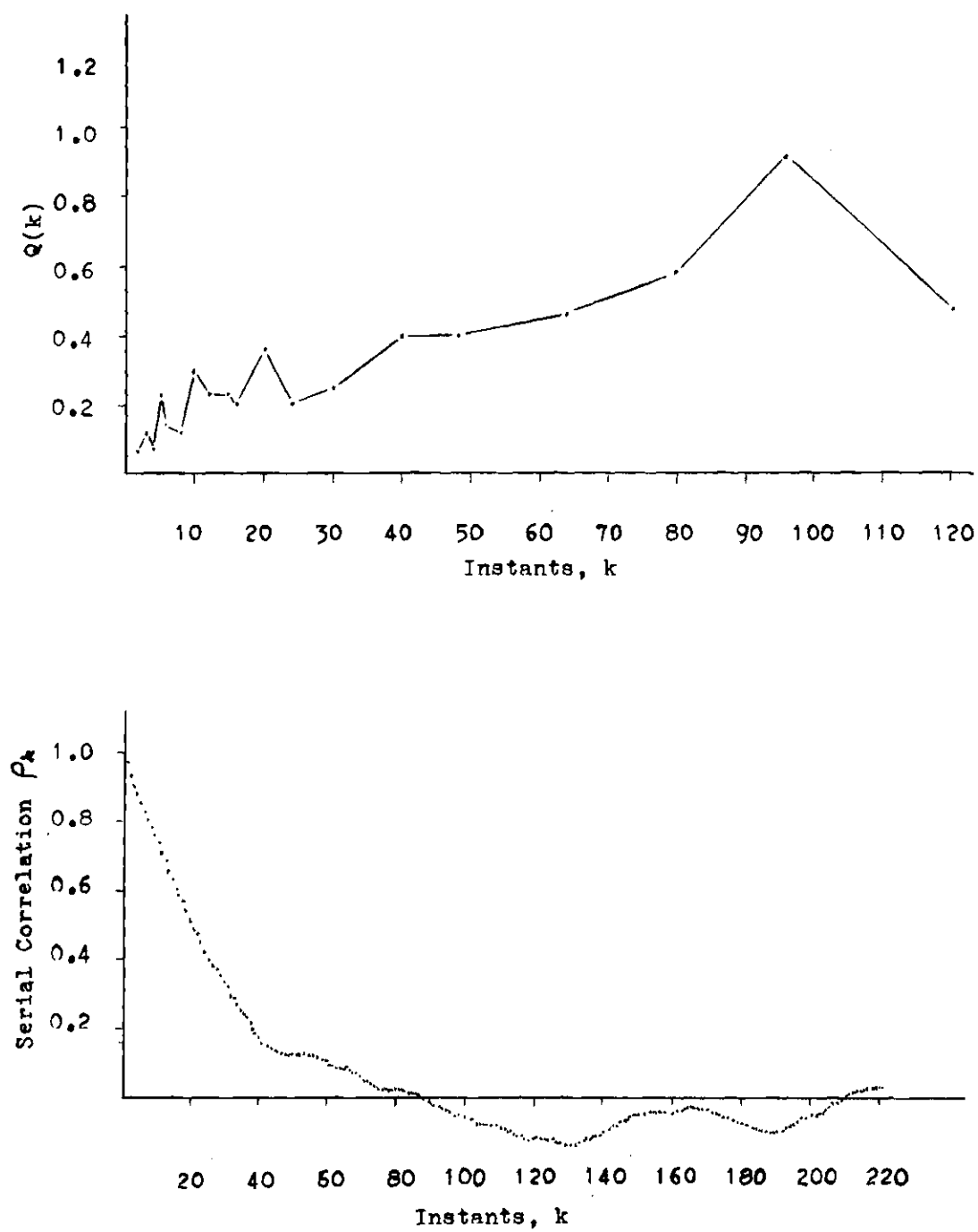


Figure 31. The Ratios $Q(k)$ and the Function ρ_k for Study No. 21

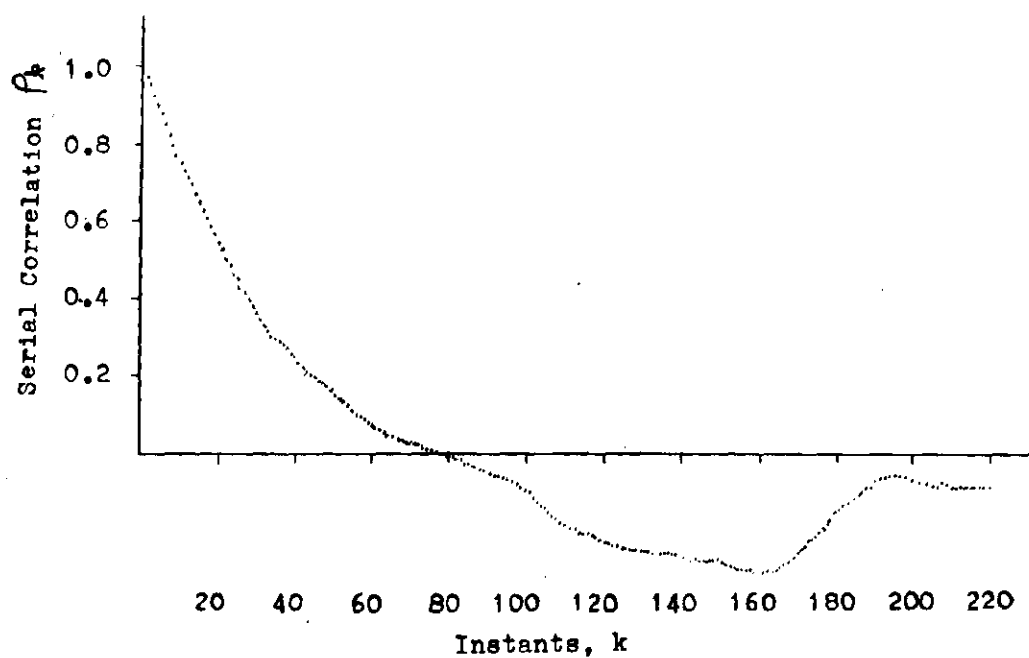
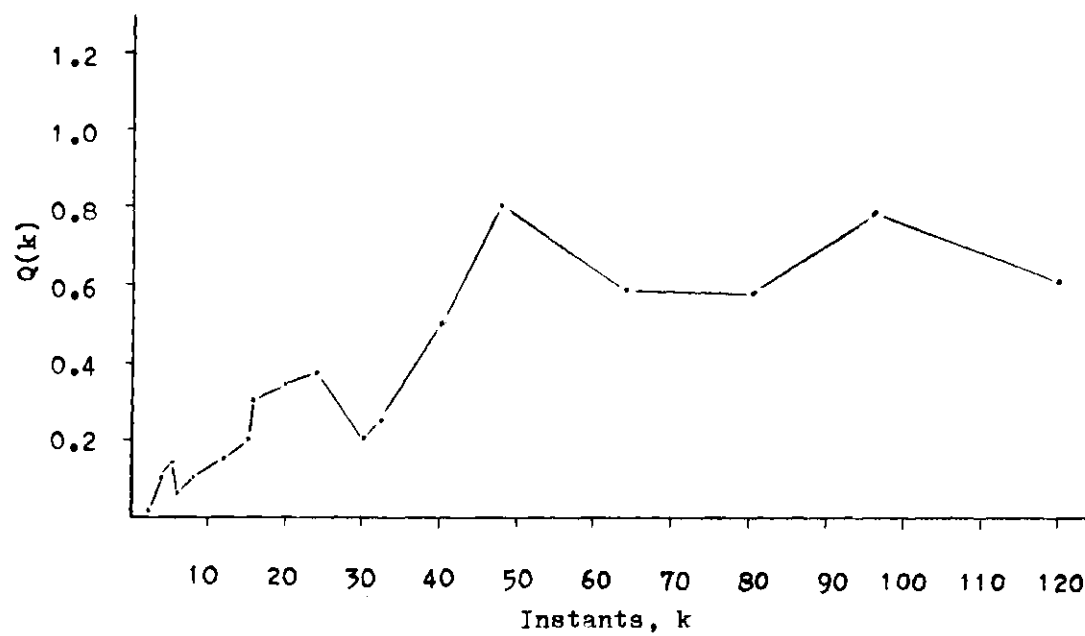


Figure 32. The Ratios $Q(k)$ and the Function ρ_k for Study No. 22

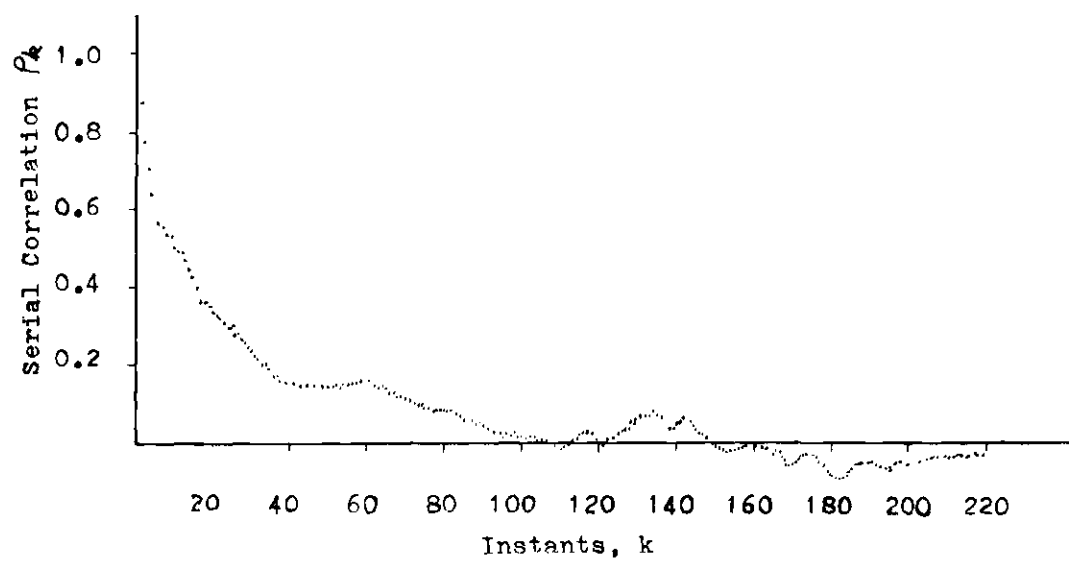
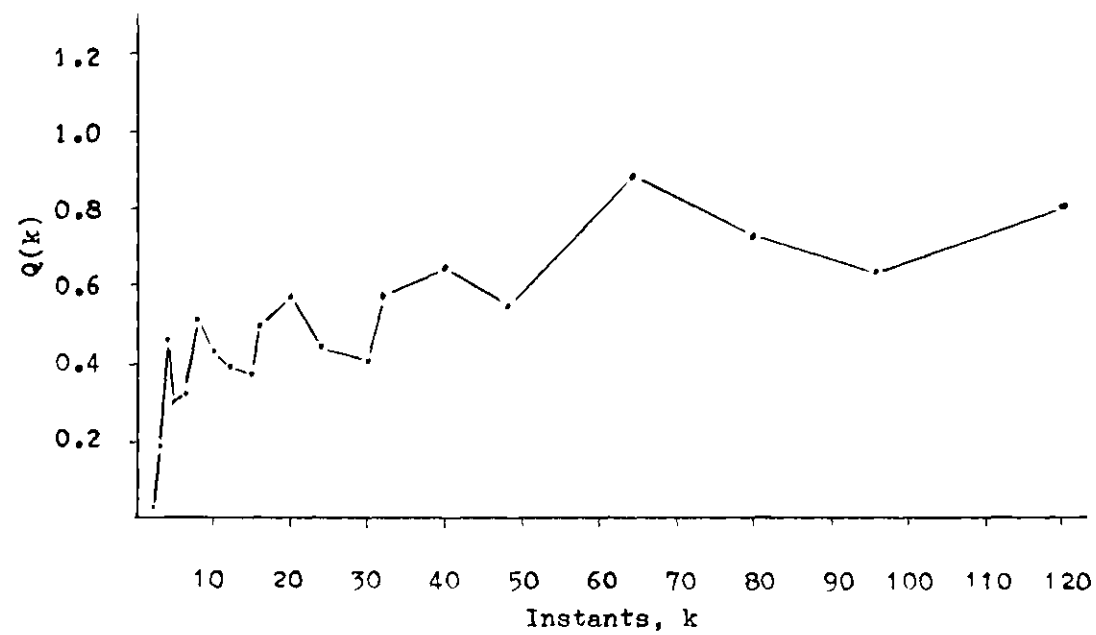


Figure 33. The Ratios $Q(k)$ and the Function ρ_k for Study No. 23

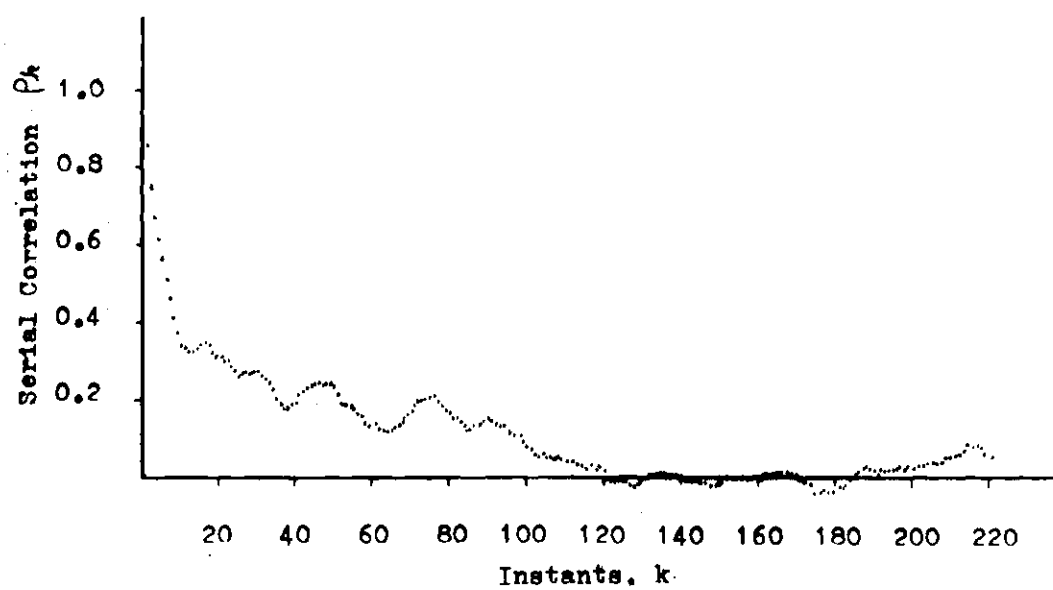
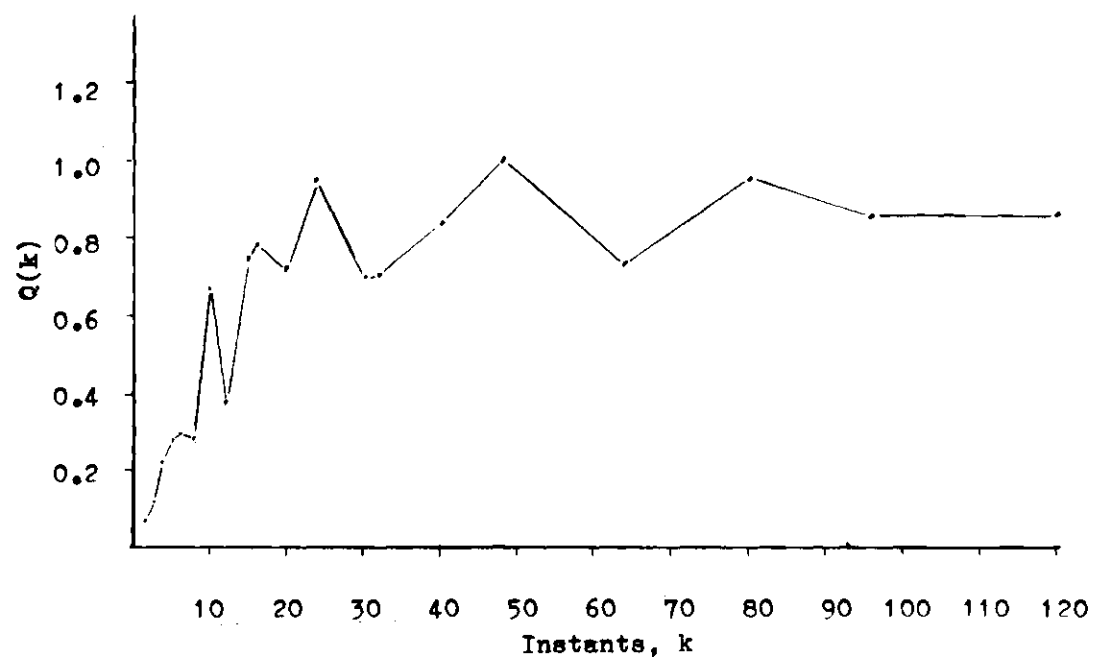


Figure 34. The Ratios $Q(k)$ and the Function ρ_k for Study No. 24

Multiple Random Starts

Study 14 was selected for testing the multiple random start theory developed in the preceding chapter. Intervals between observations were taken as 20, 64, 96, and 120 instants and there were 300 replications each of which consisted in taking ten random starts. Congruential-multiplicative methods of random generation were employed. Equations (158) and (159) were employed for the calculations, and results are presented in Figure 35. The histograms of the values $\hat{p}_{sy.k}$ are well behaved. In each case, we expect the interval

$$\bar{\hat{p}}_{sy.k} \pm 3 \bar{\sigma}_{\hat{p}_{sy.k}} \quad (168)$$

to contain 299.19 occurrences. In one case, Figure 35, one occurrence was outside the interval, and in the other three cases, the interval contained all 300 occurrences. As each estimate of p_N was calculated by Equation (158), and the variance was estimated by Equation (161), a test was performed to ascertain whether or not the confidence interval described by Equation (161) contained the true mean. This procedure was repeated for each of the 300 replications for 90, 95 and 99 per cent confidence intervals, and results are given in Table 10 for $k = 20, 64, 96, 120$.

Conclusions

Inspection of Figures 25-34 indicates that systematic sampling performed in a very satisfactory manner for the natural populations studied. One particularly significant result was the low values of $Q(k)$ for small k . For values of k less than 10, $Q(k)$ ranged from

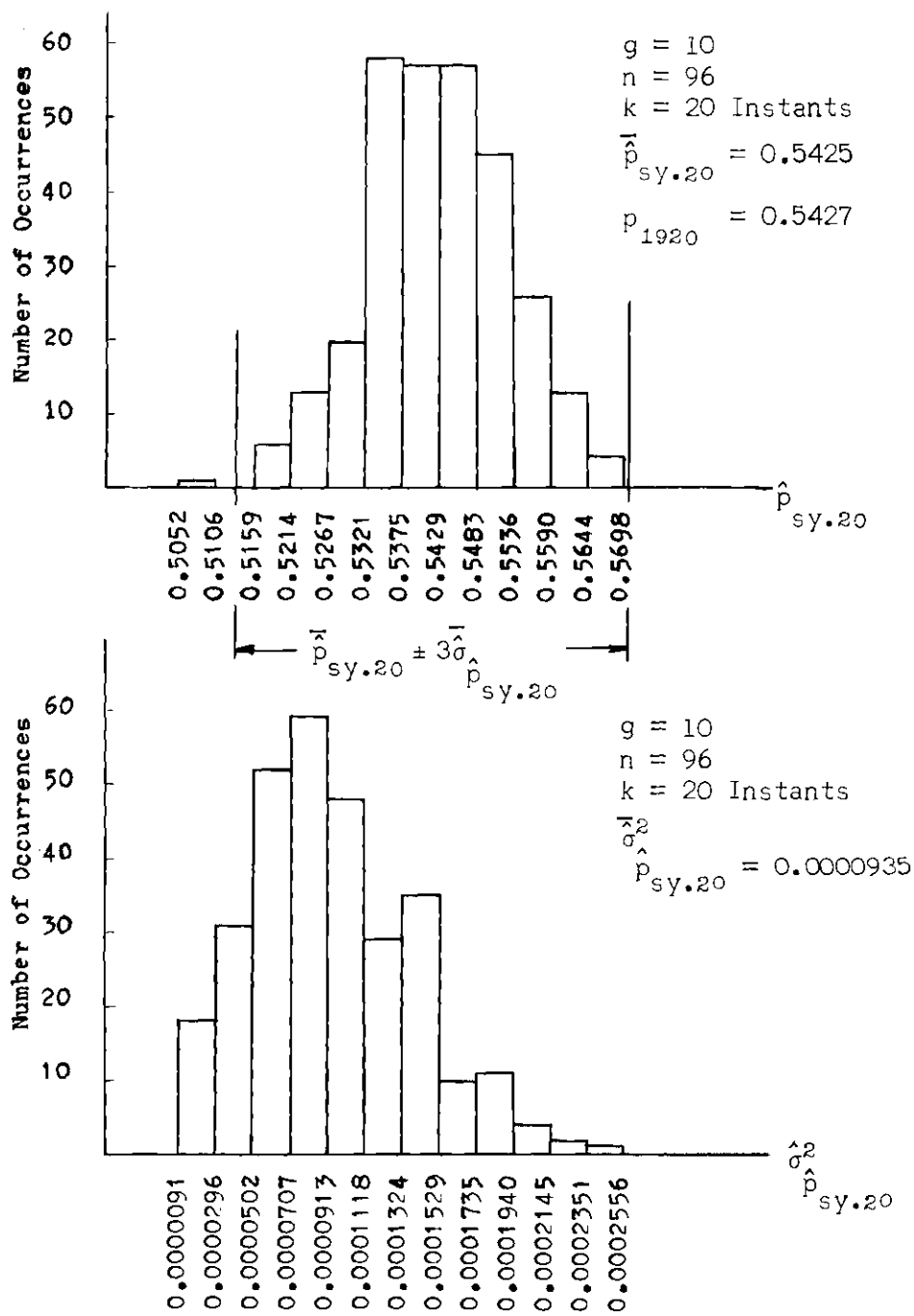


Figure 35. Results from Simulation Study of Multiple Random Starts. Study No. 14

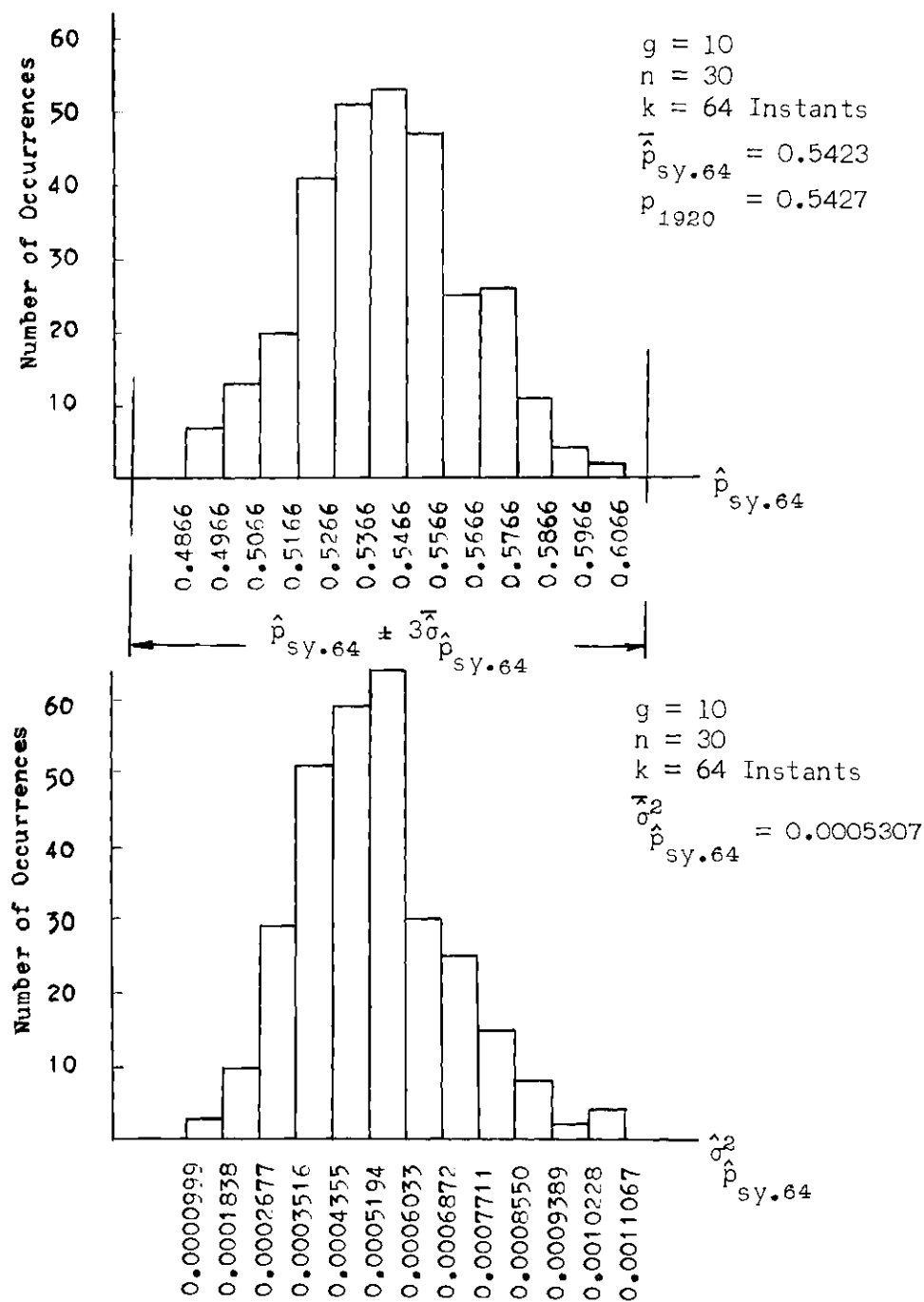


Figure 35. (Continued)

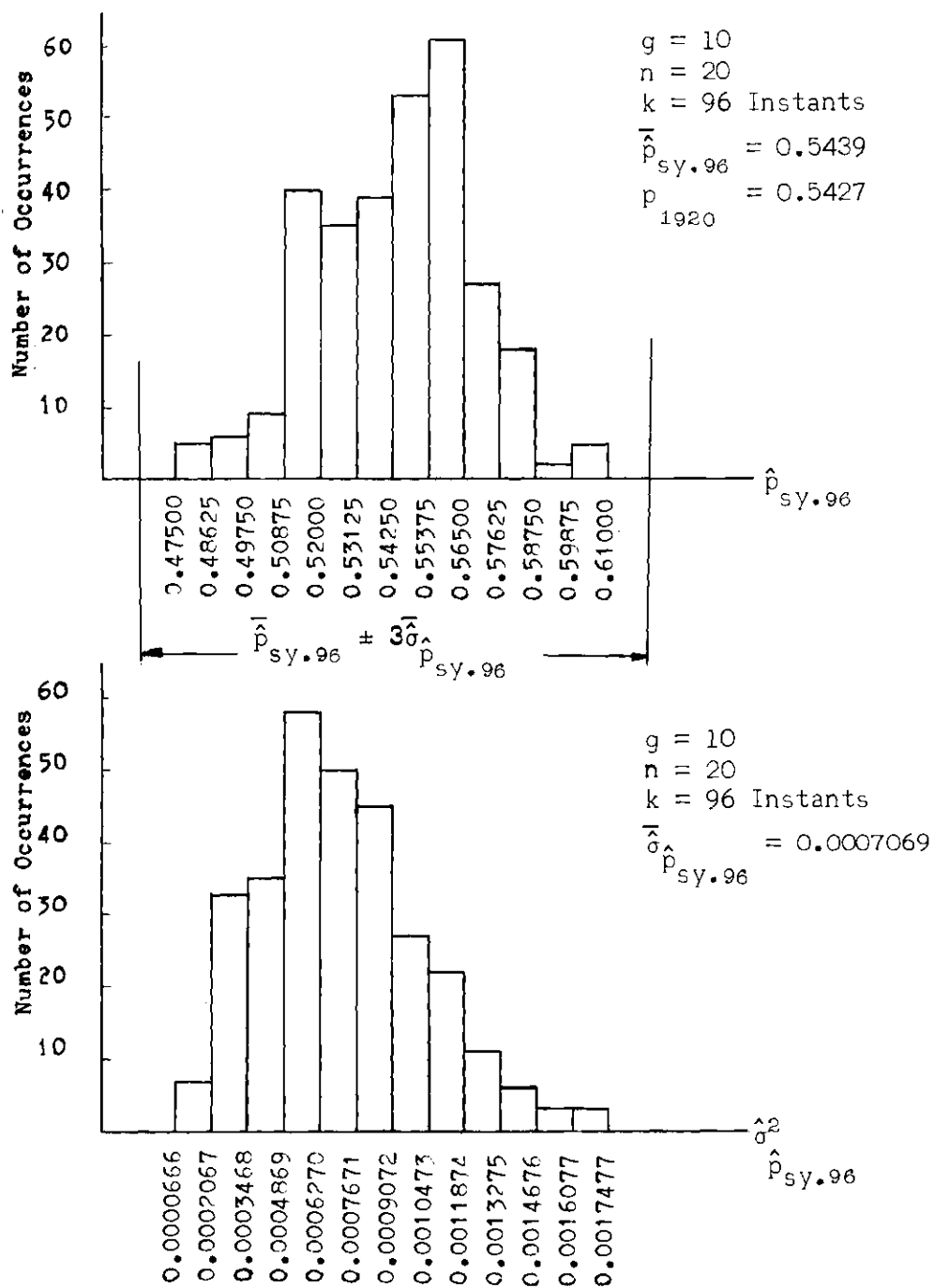


Figure 35. (Continued)

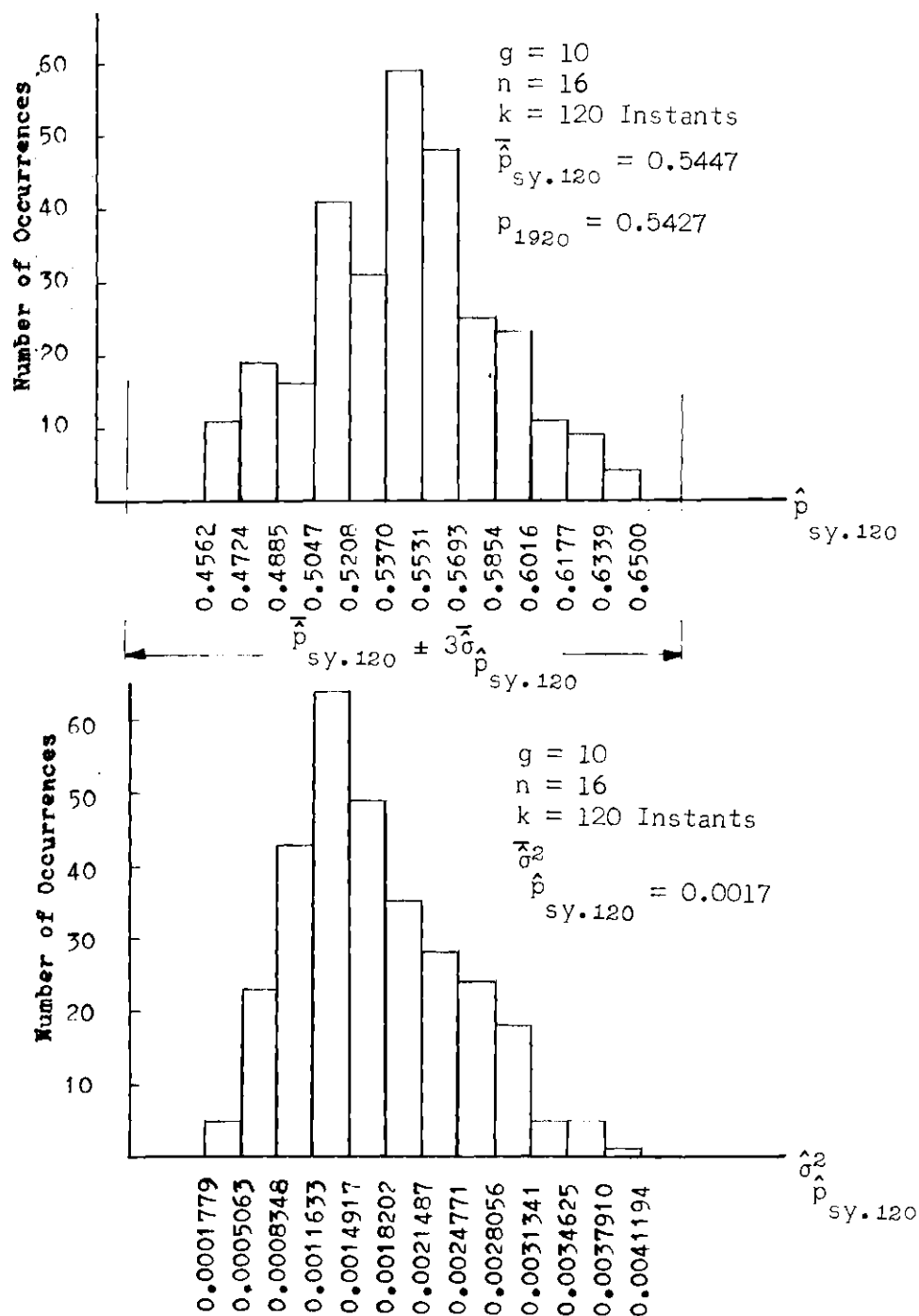


Figure 35. (Continued)

Table 10. Per Cent of Confidence Intervals
Containing the Mean

| n | k | 90 Per Cent Conf. Int. | 95 Per Cent Conf. Int. | 99 Per Cent Conf. Int. |
|----|-----|----------------------------------|--------------------------------------|--------------------------------------|
| | | $\alpha=0.10, t_{0.050,9}=1.833$ | $\alpha = 0.05, t_{0.025,9} = 2.262$ | $\alpha = 0.01, t_{0.005,9} = 2.971$ |
| 96 | 20 | 87.0% | 93.7% | 99.0% |
| 30 | 64 | 89.3% | 94.7% | 98.7% |
| 20 | 96 | 90.3% | 94.7% | 97.3% |
| 16 | 120 | 89.3% | 94.3% | 99.0% |

0.005 to 0.8. The two minute (8 instant) interval frequently employed in practice gave values of $Q(k)$ ranging from 0.10 to 0.80 with mean 0.37. Since the standard error in simple random sampling varies inversely with the square root of the number of observations, an equivalent (in standard error) simple random sample would require roughly seven times as many observations as the systematic sampling if $Q(k) = 0.37$.

Investigation of the serial correlation graphs indicates their general convex, decreasing nature, although weak periodic behavior showed up in studies 14, 22, and 24. For the larger values of k , fewer values enter into the domain of summation in Equation (75), thus a more erratic behavior would be expected for these values.

Results obtained from multiple random starts in the simulation study showed good behavior and there is close agreement between the values in Table 10 and the expected percentages, indicating the correctness in the assumption that $\hat{p}_{sy.d.i}$ is normally distributed, an

assumption required in order to construct the confidence interval described in Equation (161).

PART III

CHAPTER X

POPULATION DESCRIPTION AND SYSTEMATIC DESIGNS FOR COMPLEX ACTIVITY STRUCTURE

Introduction

The objective of this chapter is to define and describe complex activity structure. We shall define complex activity structure as the activity of m subjects, where the subjects may be either animate or inanimate. The activity of each subject may be classified as either belonging to A or A' . A pictorial model corresponding to one realization of such a process over a period $[0, T]$ is given in Figure 36. The description of complex activity structure includes the development of a quantitative model to replace the graphical model of Figure 42 and the development of quantitative measures to describe characteristics of such processes and realizations of the processes.

Infinite Population

We consider an infinite population of instants on the interval $[0, T]$, and the graphical model of Figure 36 is replaced with the following model.

$$\underline{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)] , \quad (169)$$

where

$$\begin{aligned} x_i(t) &= 1; \text{ if the activity of subject } i \text{ is in state } A \text{ at time } t. \\ &= 0; \text{ if the activity of subject } i \text{ is in state } A' \text{ at time } t. \end{aligned}$$

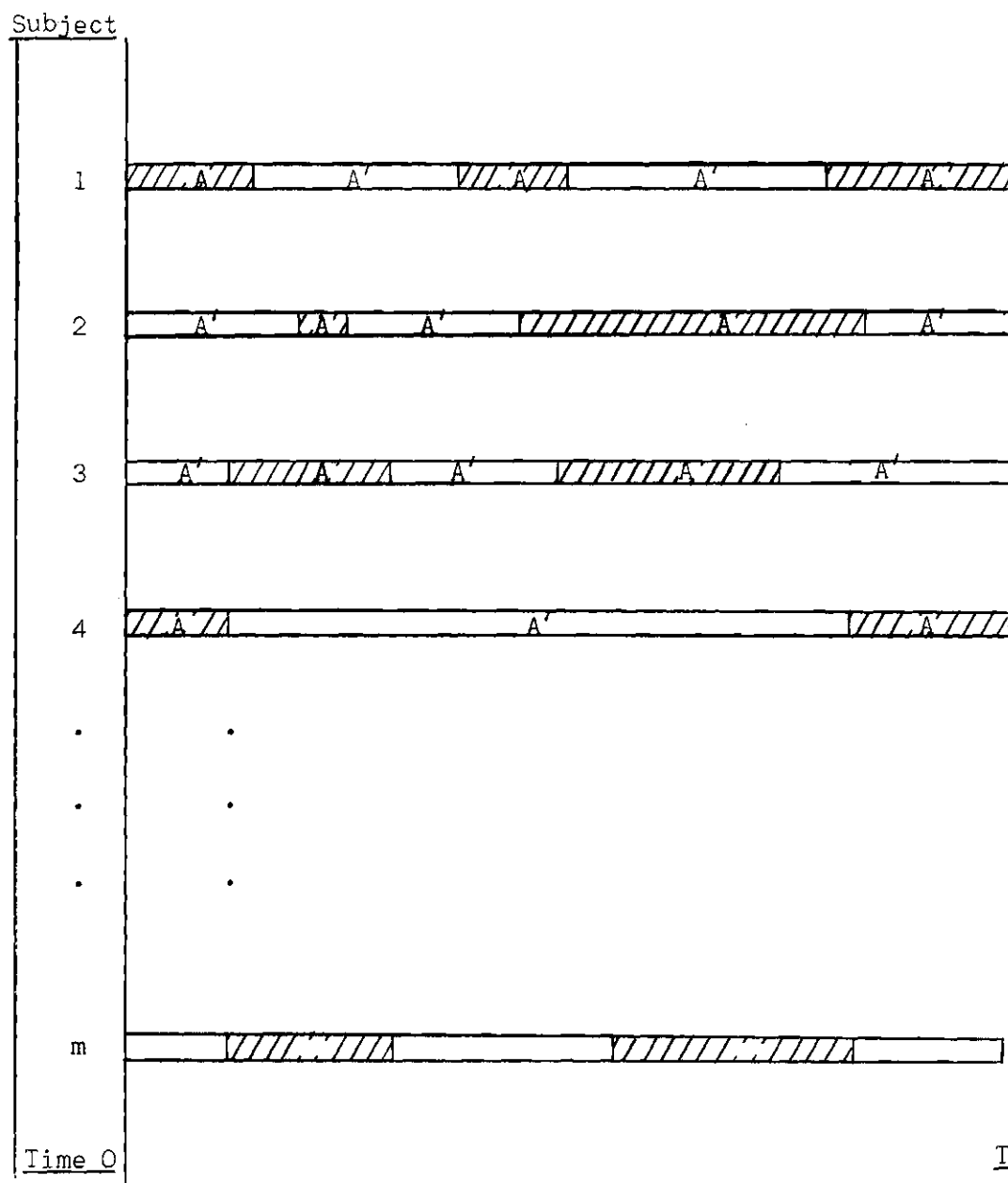


Figure 36. Complex Activity Structure

The symbol $\underline{X}(t)$ denotes the process vector of which $\underline{x}(t)$ is a particular realization.

If the inference from the sampling process is to be made about individual subjects, then the description and results of Part II apply, although several subjects are observed at the same instant. The case of interest here is the one in which inference is to be made of the collective activity of the m subjects.

If we consider the population to consist of only one realization of the process, the population mean and variance are respectively defined as

$$p_T = \frac{1}{mT} \sum_{i=1}^m \int_0^T x_i(t) dt, \quad (170)$$

and

$$\sigma_T^2 = \frac{1}{T} \int_0^T \left[\frac{\sum_{i=1}^m x_i(t)}{m} - p_T \right]^2 dt. \quad (171)$$

This population variance may be interpreted as the variance of the mean of a random sample of size one selected from the instants on $[0, T]$, where the observed variable is $p(t)$, and

$$p(t) = \frac{1}{m} \sum_{i=1}^m x_i(t). \quad (172)$$

The serial correlation between observations separated by time u is given by $\rho_{T,u}$, where

$$\rho_{T,u} = \frac{1}{p_T(1-p_T)} \left\{ \frac{1}{(T-u)} \int_0^{T-u} [p(t) - p_T][p(t+u) - p_T] dt \right\} \quad (173)$$

In the case of process characteristics rather than characteristics associated with one realization, the process is assumed to be a stationary stochastic process with mean and variance defined below.

$$p = E [P(t)] \quad (174)$$

and

$$\sigma^2 = E[(P(t) - p)^2]$$

The correlation function for points t_i, t_j is given by $\rho(t_i, t_j)$, where

$$\rho(t_i, t_j) = \frac{1}{\sigma^2} \left\{ E[(P(t_i) - p)(P(t_j) - p)] \right\}, \quad (175)$$

and if the process is covariance stationary, so that the correlation between $P(t)$ and $P(t+u)$ is a function of u only, then the correlation function is given by ρ_u , where

$$\rho_u = \frac{1}{\sigma^2} \left\{ E[(P(t) - p)(P(t+u) - p)] \right\}. \quad (176)$$

Finite Population

The argument for consideration of finite populations has already been expressed, and we shall continue to consider a time period with $N = nk$ instants, where N, n , and k are integers. Associated with the j^{th} instant there is a variable vector \underline{X}_j and a particular realization of \underline{X}_j is denoted by \underline{x}_j , where

$$\underline{x}_j = [x_{1,j}, x_{2,j}, \dots, x_{m,j}] \quad (177)$$

If we consider a particular finite population of one realization, the mean, variance and serial correlation are shown below, where

$$p_j = \frac{1}{m} \sum_{i=1}^m x_{i,j} \quad (178)$$

$$p_N = \frac{1}{nk} \sum_{j=1}^{nk} p_j \quad (179)$$

$$\sigma_N^2 = \frac{\sum_{j=1}^{nk} (p_j - p_N)^2}{nk}$$

$$\rho_{\delta k} = \frac{1}{\sigma_N^2} \cdot \frac{1}{k(n-k)} \sum_{j=1}^{k(n-\delta)} (p_j - p_N)(p_{j+\delta k} - p_N)$$

where δ is an integer.

If we consider the variable and not just one realization, and assume a covariance stationary stochastic process, then the process mean, variance, and serial correlation are respectively given below.

$$P = E[p_j] \quad (180)$$

$$\sigma^2 = E[(p_j - p)^2]$$

$$\rho_u = \frac{1}{\sigma^2} \left\{ E[(p_j - p)(p_{j+u} - p)] \right\}$$

Systematic Sampling Designs

Both single random starts and multiple random starts are considered as was the case in Part II. The designs are the same as those considered earlier except that the observed variable is no longer a zero-one variable. The finite population concept is to be employed, and p_j is observed in the case where the j^{th} instant is a member of the sample.

CHAPTER XI

ESTIMATION OF PARAMETERS ASSOCIATED WITH COMPLEX ACTIVITY

Introduction

A detailed analysis of the properties of estimates associated with complex activity structure is not within the scope of this research, and it is the objective of this chapter to briefly describe such estimates and indicate their properties in situations where the properties are obvious. The finite population concept is employed in order to make use of the existing literature of cluster sampling.

There are numerous situations in practice, as evidenced by the literature of activity sampling, in which the observer observes several subjects and estimates the fraction of total time spent in a given activity classification or state.

Single Random Start

In the case of a single random start, one of the instants $1, 2, 3, \dots, k$ is randomly selected, and every k^{th} consecutive instant thereafter is selected. The observation associated with each observation is a number on $[0, 1]$, and for the j^{th} observation, this number is denoted by p_j .

Considering a particular realization of the process, and letting y_{ij} represent the value of the j^{th} observation in the i^{th} possible systematic sample, so that the p_j values replace the x_j values of Table 7, the analysis of Equations (141 - 147) may be employed. Equation (147) denotes the condition that is both necessary and sufficient for the estimate

$$\hat{p}_N = \frac{1}{n} \sum_{r=0}^{n-1} p_{w+rk} \quad (181)$$

obtained from the systematic sample of n instants to be more precise than the corresponding estimate obtained from a simple random sample. In the above expression w represents the first instant randomly selected from the instants $1, 2, 3, \dots, k$.

This estimate is unbiased due to the random start; however there is no method of obtaining an unbiased estimate of the variance of the estimate. In the case where the inequality of Equation (147) is satisfied, an upper bound to the variance of the estimate may be expressed by the following relation:

$$\hat{\sigma}_{\hat{p}_N}^2_{UB} = \frac{1}{n(n-1)} \sum_{r=0}^{n-1} (p_{w+rk} - \hat{p}_N)^2. \quad (182)$$

In the case where the study extends over more than 20 periods or strata, and the parameter to be estimated is p_{LT} defined as

$$p_{LT} = \frac{1}{L} \sum_{h=1}^L p_{N \cdot h}, \quad (183)$$

where L is the number of strata and $p_{N \cdot h}$ is the value of the parameter in stratum h , then the estimate of p_{LT} is given by Equation (153) if $\hat{p}_{sy.d.h}$ is replaced by $\hat{p}_{N \cdot h}$. The results of Equation (154) apply as an overestimate of the variance of \hat{p}_{LT} , and the approximate confidence statement of Equation (156) is applicable.

If the process variable is considered rather than a particular realization, then Cochran's Theorem giving sufficient conditions for the superiority of systematic sampling may be employed in cases where applicable, and an upper bound on the variance is given by Equation (182). The difficulty in these cases arises in ascertaining cases in which the theorem is applicable, and further research is needed in this area.

Multiple Random Starts

In the case of multiple random starts much of the estimation difficulty is overcome, and the results of Chapter VIII may be directly applied with modification to the extent that the observed variable is a fraction rather than a zero-one variable. At the present state of research, multiple random starts are recommended for the study of complex activity structure.

Concluding Remarks

In the cases where the inferences resulting from the sampling is to be made about individual subjects rather than the group of subjects, the observational vector may be divided into its components, and the analysis of Part II of this thesis is applicable. In the case where the activity of the group is of interest, the cluster sampling concept is utilized and the variable measured is a ratio rather than a zero-one variable. With this modification, the results of Part II may be employed in some cases where sampling is to extend over several strata and/or multiple random starts are employed. The difficulty arises in ascertaining situations in which systematic sampling is as precise as random sampling. The concept of cycle length that was employed in Part II is no longer applicable.

In the situation where the activity of the subjects is independent, intuitive reasoning leads to the conclusion that systematic sampling is a good design since periodic effects of individual subjects tend to cancel out when the cluster is observed. This statement is made as an untested hypothesis, and further research is recommended on this aspect of systematic sampling.

CHAPTER XII

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

There are various complex sampling designs in which systematic sampling may be employed in the collection of data. As stated in Chapter I of this thesis, only simple activity structure is treated in detail herein, and there is some work yet to be done in the development of complex designs employing systematic sampling at some stage. In the cases where stratified procedures were developed, strata of equal sizes were considered since it is felt that this is the case most usually encountered by the activity sampling practitioner, when shifts or half-shifts form the strata.

In the case of the simple activity structure, there are numerous situations in which systematic sampling yields an estimate having less variance than simple random sampling. Analytical and Monte-Carlo techniques were employed in this investigation of this structure, and two special cases as well as the general case were considered.

In the first special case, the situation was considered in which the activity of interest occurs at most once on each time period. In this case, systematic sampling is uniformly more precise than simple random sampling.

The variances of the estimates from simple random and systematic sampling are compared in Figure 4 and Equations (88) and (89). An upper bound to the variance associated with systematic sampling is given in

Equation (90). The distribution of the estimate associated with the systematic design was analytically determined and validated by simulation employing the Monte-Carlo method. A powerful confidence statement results from this determination, and this statement is given in Equation (91). Stratification techniques are utilized where the study is to extend over several time periods; and the estimate of the parameter, an upper bound to its variance, and a confidence statement on the parameter are given in Equations (94), (95) and (96) respectively. The methods of Yates (164) are employed in the development leading to Equations (81), (83) and (85), and these results agree with Yates results; however, all subsequent developments were obtained herein.

In the second special case, the situation was considered in which the interval between observations is shorter than all occurrences of both A and A'. This case was presented in 1960 by Davidson, Hines, and Newberry (47); however, the assumptions of the earlier work were not altogether correct, and these are investigated herein. The estimate and an upper bound to the variance are given in Equation (119), and non-parametric methods are employed in the statement of confidence in Equation (120). Stratification techniques were again employed for cases where study extends over several time periods. The estimate, an upper bound to its variance, and a confidence statement on the parameter for this situation are given in Equations (127), (123) and (124). The comparison of simple random and systematic designs is made in the concluding remarks to Chapter VI. Systematic sampling is the superior design for most cases of this type encountered in practice, and this development is particularly applicable to the utilization of photographic techniques for the collection of data.

In the general case, where the assumptions of the two special cases may not be employed, the distributions of span lengths for both A and A' were considered. Cochran's theorem (38, p. 117) on systematic sampling was employed in showing that exponential distributions of these span lengths gives rise to a correlation function having sufficient properties to assure the superiority of systematic sampling. Monte-Carlo simulation methods were employed in the investigation of the correlation function when the span lengths have gamma and truncated normal densities. The resulting correlation functions are given in Figures 7 - 24. The functions are not monotone decreasing convex, and the variance of the estimate is dependent upon the relationship between the interval between observations and the mean cycle length. The factors affecting the form of the correlation function are the coefficient of variation in cycle length and the ratio of the expected values of the span lengths of A and A' . The analysis of the data resulting from the simulation studies indicates that in most practical applications the practitioner may expect systematic sampling to yield an estimate having less variance than that obtainable from simple random sampling when the interval between observations is small or very large relative to the mean cycle length. Specific values and more detailed conclusions are presented at the end of Chapter VII.

Estimation procedures are developed for the estimation of variance in the general case. These results are presented in Equations (150), (154), (159), and (162). Designs employing one random start and multiple random starts are considered, and confidence statements are developed and presented in Equations (156), (161), and (164). Some other specific conclusions are stated at the end of Chapter VIII.

The methods and formulae developed herein for simple activity structure were tested on ten natural populations of human activity and systematic sampling performed very well in each case. The results are presented graphically in Figure 25-34, and specific conclusions are presented at the end of Chapter IX.

A brief investigation of complex activity indicated the applicability of estimation procedures developed in connection with simple activity with the modification of considering the variable measured as a fraction rather than a zero-one variable.

Recommendations

The recommendations for additional studies may be classified into two groups, theoretical and applied.

Theoretical Studies

1. Studies should be undertaken to resolve problems associated with complex, multi-stage, sampling designs where systematic sampling is employed in the collection of data.
2. Cluster sampling and estimation techniques, where clusters are of unequal sizes, should be developed.
3. Standard procedures for estimating period-to-period variability should be developed.

Practical Studies

Since the ten natural populations studied herein consist of activity of an irregular nature, additional studies should employ the theoretical structure developed here for the study of some short cycle, repetitive activities.

A P P E N D I X A

APPENDIX A

A REVIEW OF THE HISTORICAL DEVELOPMENT OF ACTIVITY SAMPLING

Introduction

This appendix reviews and criticizes the historical development of activity sampling with emphasis on the papers not discussed in chapter II. The literature may be divided into two broadly defined groups as that which has been published in technical journals and proceedings of professional conferences, and that which has been published in trade periodicals. In general, the technical journals and conference proceedings have published manuscripts which contributed to the basic knowledge of the application of sampling techniques to activity analysis as well as case studies some of which report methodological innovations. The quality of presentation in the technical journals has been superior in terms of scholarly development, technical content, and mathematical correctness.

In this Appendix, the literature is reviewed as it applies to research, and a "significant contribution" is considered to be an extension of the frontiers of knowledge at the time of publication. Many of the papers and articles considered insignificant herein were, no doubt, of considerable value to the practitioners who needed stimulation and methodologies for data collection and analysis as well as confidence in statistical measurement. A statement that a particular presentation was insignificant does not imply that the presentation was not of value to the community of practitioners.

The body of knowledge which has been defined as activity sampling

was an outgrowth of the scientific management movement of the late nineteenth and early twentieth century. Although some time elapsed before statistical sampling techniques were introduced, the concept of activity analysis directed towards measuring worker or machine idle time had earlier received the attention of the leaders of the scientific management movement.

The Introduction of Activity Sampling

In 1935, L. H. C. Tippett (148) reported the application of statistical methods to several measurement problems arising in textile research. His interest was confined to attribute type measurement, and his presentation dealt with applications of the binomial and Poisson probability distributions. By assuming small sample size relative to population size in the case of finite population sampling, the necessity for considering the appropriate hypergeometric distribution was circumvented.

In February, 1935, the second section of Tippett's work introducing activity sampling was published. This presentation was entitled "A Snap-Reading Method of Making Time-Studies of Machines and Operatives in Factory Surveys," and the term "snap-reading" was applied to activity sampling from this time until 1941. The following paragraphs, quoted from the introduction to this section of his work reveal Tippett's motivation and objectives:

Investigation into problems of management often involves a knowledge of the rate of output of various machines, of the durations of machine stoppages, or of the way in which the operatives distribute their time. For example, when several kinds of cloth are being woven at the same time in one mill, the correct proportion of overhead costs that each should bear can be estimated only if the relative rates of output are known. Attempts to reorganize the work of the operatives can be more appropriately made if there is a full knowledge of how much of their time is spent in

performing their various duties, and how much is available for supervision or relaxation. A rational attempt at increasing the output of any machine can only be made if the amount of productive capacity lost for each of various causes is known. Indeed, without data of the kind referred to, "scientific management" is scarcely possible.

Methods exist for obtaining the kind of information under discussion, but they have their difficulties and limitations. Thus, for measuring the output of most types of machines, indicators may be fitted and read periodically, but this may sometimes be expensive and inconvenient if it has to be done specially for an investigation.

If an analysis of causes and durations of machine stoppages is desired, an observer may make a time study on a few machines, using either a stop watch or a mechanism for recording automatically the durations of the stoppages. This method is suitable in a weaving investigation if only a few looms are involved, but it would take a long time to obtain a reliable average for a whole shed since it is only possible to observe eight looms or fewer at a time. Further, because of the continued presence of the observer, the weavers may behave abnormally, and spurious results may be obtained. Similar difficulties arise when observing the machines in other processes.

If the time spent by the operatives in various parts of their duties is required, continuous observation with the aid of a stop watch is usually resorted to, and the method is attended with the same difficulties as those mentioned above, in addition to the personal objection of operatives to being timed. Also, when the operative has a large number of short miscellaneous duties and frequently changes his actions, as in many textile processes, timing becomes very difficult.

In spite of these difficulties, the more usual methods of observation yield reliable results when applied with care, and they may have to be used as standards by which the accuracy of other methods is judged.

This paper describes an alternative method, which consists in taking a large number of snap readings of the machines at random instants, recording their state (i.e., whether working, or, if stopped, the cause of stoppage). The percentage of observations recording the machines in a given state is a measure of the average percentage of time they are in that state. These data, supplemented by observations of speeds and rates of occurrence of various causes of stoppage, give the information required of a time study. Similar snap readings may be taken recording what the operatives are doing (148, pp. 51-52).

Although Tippett mentioned random instants for observations, his studies did not employ any such scheme. The studies were based on a stratified sampling method, where days served as the basis for stratification, and the instants within a stratum were observed at "convenient" intervals.

Errors in sample results were presented in detail. Systematic errors leading to the bias of the estimate were presented, and Tippett suggested elimination of this bias by making the observation a "snap reading," and gaining confidence of the operatives in the case where the activity is manual or manually controlled to a high degree. Random variation is described by the binomial model where the

$$\text{Standard Error of } \hat{p} = \sqrt{\frac{p(1-p)}{n}},^*$$

and \hat{p} is the estimate of the fraction of time, p , that the activity is in the state of interest, and n is the number of observations. The effect of stratification is not considered, and in cases where two or more observations found the activity in the same state on the same occurrence of the state, only the result of the first observation was recorded. Process variation is recognized as man-to-man variation, machine-to-machine variation, and time-to-time variation. The time-to-time variation is broken down into day-to-day variation and period-to-period variation within the day.

Several studies were presented, and Pearson's "goodness of fit" test

* Private correspondence with Mr. Tippett, September 5, 1962, "I now think that in my 1935 paper I was wrong.... I did not for reasons of convenience, randomize the observations; but non-randomization is likely to give a bias only if there are coincident periodicities."

was employed for testing for conformity to the binomial model. Replication was over time resulting in larger variability than explained by the binomial model. Analysis of variance techniques were employed to show the significance of the day to day variability.

This excellent work received no appreciable attention for six years; however, studies were continued by the Shirley Institute.*

Early Activity Sampling Studies in the United States

In December of 1940, Morrow (108) reported results obtained in "practical tests in three separate industries by graduate time study students of New York University" (108, p. 302). The students were Lynch, Isaacson and Schaeffer. Morrow called Tippet's sampling technique "Ratio-Delay," a name which followed activity sampling from this time until 1952. He summarized Tippet's work, adding nothing but his approval. The "practical tests" are not described in the published abstract, and the following twelve conclusions are given:

1. Only homogeneous groups should be combined; such as delays on similar operations performed on similar types of machines, or delays of operators on work of a similar nature.
2. A large number of observations is recommended, and studies are best adapted to large groups of machines or operators. When the number of observations on the job was about 500, a fairly reliable result was obtained. Over 3000 observations gave very accurate results.
3. Results from a few hundred observations may be used, if the frequency distribution conforms to the binomial law.
4. The accuracy of results may be determined in any case.
5. As the percentage of delay time increases, more observations are necessary for a given accuracy.

* The British Inter-Industry Textile Research Organization.

6. Data are more reliable if the observations are taken over a long period of time.
7. Observations must be taken at random intervals and distributed over all hours of the day and week.
8. Intervals between samples must be sufficient to give independent readings.
9. The ratio delay study provides an opportunity to observe and evaluate operations of the department as a whole.
10. The observer's work may be interrupted at any time without affecting the study. Taking studies is not tedious for the observer.
11. There was no objection to ratio delay studies by operators, because no stop watch is used and the operators are not closely observed for a long period.
12. The cost of studies is about one third that of production studies (108, p. 303).

Although Morrow's paper indicated a lack of theoretical understanding of the subject on which he was reporting, his work was significant in that it was the introduction of activity sampling to the United States.

In July of 1941, Schaeffer (142) reported a study conducted at the J. E. Ogden Company where activity sampling was employed to determine machine idle time. The sampling was stratified by day with thirty observation trips per day. There was no attempt to treat the problem concerning the goodness of the estimate, and the method of selecting the "random intervals" for observation was not given.

In their Production Handbook, published in 1944, Alford and Bangs (9) summarize Morrow's work. There is no explanation of the method for selecting instants for observation or of the precision of the resulting estimate, and Morrow's twelve misleading conclusions are restated. It is note-worthy that this handbook was recommended to students of industrial engineering for twelve or more years after its publication. The Handbook

of Industrial Engineering and Management edited by Ireason and Grant and published in 1955 (76) gives a brief presentation of activity sampling that is no more complete than the Alford and Bangs work; however, the misleading statements are not included.

Morrow authored a text (109) published in 1946 in which he devoted twenty-five pages to activity sampling which was recommended for determining delay allowances to be used in connection with stop-watch time study. The case studies which Morrow had presented to the A.S.M.E. group in 1940 constitute the major part of this material, and this is not a significant contribution to the knowledge of activity sampling.

In 1947, Tippet (149) read a paper before the Royal Statistical Society in which he described the work of The Shirley Institute related to industrial productivity and efficiency. A large part of this paper was devoted to a review of activity sampling and measurement objectives. In connection with the description of activity sampling, the following statement regarding observation instants is given:

If any machines are stopped for a particular cause for an average of x per cent of their time, and a large number of snapreadings of the states of the machines are taken at intervals uncorrelated with the occurrence of stops, the percentage of readings recording the machine as stopped for that cause will tend to be x (149, p. 109).

Tippet does not recommend random time observations in this paper, but he recommends observation at instants uncorrelated with machine stoppage. There are no suggestions as to how such instants might be selected.

In April, 1948, Abruzzi (2) recommended the control chart for ascertaining statistical control or stability of the parameter, p , being estimated by activity sampling study. Stratification by day is suggested; however, other bases are considered. Abruzzi suggests that the chief

advantage of the control chart method is that it is the most economical and effective method of ascertaining stability over time, estimating the parameter of the binomial model, and detecting assignable causes of variation which provide effective guides to corrective action. A second paper appeared in May, 1948 (3) in which Abruzzi presented some case histories of control chart applications. Stratification is used with time being the criterion for the formation of strata. The strata are of equal size of one and one-half hours in the first study and unequal sizes from one to two hours in the second study. Instants were randomly sampled within the strata. Although stratified sampling is employed, Abruzzi calculates the variance of the estimate using the model for simple random sampling, i.e.

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n},$$

and he seems unaware that the stratified sampling technique he utilized will generally reduce this variability. Advantages afforded by a preliminary study are listed, and results of control chart analyses are presented. After stability was ascertained the analysis of variance technique was employed to test for the significance of the variation between strata within days. In these cases, this variability was not significant.

In October 1948, Petro (126) reported a case study at Thompson Products which utilized the activity sampling technique to establish allowances for time standards; and in November, 1948 Abruzzi (4) reported several useful guides for using control charts on productivity indices. Neither of these works is particularly significant.

Activity Sampling 1950-1959

In August, 1950, Correll and Barnes (43) summarized the work of

Tippett and Morrow. The work of Abruzzi was not mentioned, and the general question of stability was not considered. Inference was to be about the time period over which observations were made. Eight "precautions" are suggested of which two are most indicative of the technical weakness in this presentation. They are:

Observations must be taken at truly random intervals....
Sufficient time must be allowed between rounds so that same delay is not recorded twice... Long delays should be only recorded once as the Ratio-Delay theory is based on the percentage number of delays rather than their length (43, p. 12).

Three case studies are presented, and six general advantages are claimed for Ratio-Delay as follows:

- (1) Less time is consumed in determining the allowances, especially where there are many similar operations so that several observations may be obtained each round. The estimated cost of determining allowances by Ratio-Delay is about one-fourth to one-half that of time study.
- (2) There is less chance of obtaining misleading results as the operators are not under close observation for long periods of time.
- (3) Observations may be taken over a period of days or weeks thus decreasing the chance of day to day or week to week variation affecting the results.
- (4) This type of study may be interrupted at any time without affecting the results.
- (5) The Ratio-Delay method may be used to analyze an entire department for the purpose of methods improvement.
- (6) It is possible by this method to obtain delay allowances on operations where time-study is not feasible.... (43, Part II, p. 17).

In enumerating the practical problems encountered, Correll and Barnes make two points which take their work out of the realm of science.

Care had to be exercised in obtaining a representative sample. If an unusual proportion of the observations were taken at the beginning or end of the day, the delay percentage might be higher than should be expected.... The determination of the

time interval between observations also posed a problem. Enough time must elapse to assure the observer that independent readings are being obtained. A practical expedient employed was the spacing of intervals so that they were slightly greater than the maximum time-length of normal delays (43, p. 18).

This article was the first of Barnes' publications on this topic, and it seems to have had a considerable influence of practicing engineers in the industrial community.

During the Industrial Engineering Institute Conference in 1951, an annual conference at the Los Angeles branch of the University of California, Sammet (139) presented research findings where results from activity sampling studies are compared with results obtained from concurrent continuous production study. Sammet does not claim "truly random" observation times, and he does not discuss his method for selecting times for observation trips through the work area. He advocates recording results from all observations which is the first reported departure from Tippett's method of recording long delays on the first observation only. This presentation is oriented towards determining delay allowances to be used in establishing time standards for manual work. The question of stability is not explicitly treated, and the comparison is not made on a statistical basis. Differences may or may not be significant.

Allderige (10) presented a paper in September 1951 in which the economies of sampling for measurement are emphasized. He advocated control chart application to ascertain stability and nomography for sample size calculations. This paper does not report experimental evidence or new analytical development.

Sammett and Hassler (140) published an expanded form of the paper given by Sammett earlier in 1951. Sequential sample size calculations

are developed in this later publication.

Brisley (31) presented a paper in 1951 before the Society for the Advancement of Management group in New York. This was the first of several works presented by Brisley and like the following work, the 1951 presentation was a persuasive type, non-technical paper, and no new theories or methods were incorporated. Cogan and Stilliam (40), industrial engineers with Warner-Hudut, Incorporated, presented a testimonial for "Ratio-Delay" in Factory Magazine.

In 1952, the Editor of Factory, H. L. Waddell (154), suggested the name "Work Sampling" be used for activity sampling since the term "Ratio-Delay" implied use of sampling for delay study only. This new name became popular at this time, and it has followed activity sampling for eleven years. Waddell's editorial was accompanied by an article authored by Brisley (30) which followed the pattern of the rest of Brisley's publications. In his sales oriented testimonial Brisley claimed that work sampling:

Gets the facts at one third to one sixth the cost of continuous observation.

Does not require observers with special skill and training.

Provides the accuracy required.

Makes it practical to get facts that you wouldn't otherwise try to collect.

Produces fewer complaints than continuous observation from the operators under study.

Produces less distortion than continuous observation in the operator's normal work routine (30, p. 85).

Nadler and Denholm (114) presented a paper in 1952 in which they advocated control chart application to ascertain stability and lack of

process control. Although Abruzzi's work had covered the same topic in greater detail, this paper was significant in that it brought Abruzzi's work to a more understandable level for industrial personnel. Bogenrief (27) presented a testimonial for work sampling in December, 1952; however, this contribution was insignificant.

Abruzzi authored a book (5), published in 1952, in which he criticized the work of Tippett on the basis of the inconsideration of statistical stability. The text represents Abruzzi's papers which advocated the control chart applications (2, 3, 4).

Davidson (45) presented a monograph which was sponsored and published by the American Institute of Industrial Engineers in 1952. In this work, he critically reviewed time study practice stating that Tippett's work had several characteristics which distinguished it from other work in time study. These distinguishing characteristics are noted as:

1. The technique rests on a carefully considered theoretical foundation with basic assumptions clearly stated.
2. The sources of errors in results obtained from the technique have been investigated theoretically and empirically and have been fully discussed.
3. The technique provides a method of computing limits of accuracy for the result together with a statement of the probability that those limits may be exceeded.
4. The technique includes an ingenious criterion (which does not depend upon subjective judgments) for accepting the results of a particular application or rejecting them. Rejection may arise from:
 - (a) inapplicability of the assumptions on which the method is based,
 - (b) Errors in observations, or
 - (c) insufficient number of observations.
5. The technique was subjected to objective validation before it was publicly announced (57, pp. 64-65).

McAllister and Nelson (102) presented a criticism of Tippet's 1935 paper in which they review his work, and take issue with the "elimination of very long delays from consideration" between repeated readings on the same machine (102, p. 2). It is shown that a biased estimate results from the procedures quoted above which had been advocated in most of the literature prior to this time.

Gaylord (61) presented a paper before the 1953 Industrial Engineering Conference, held at the West Virginia University. Two applications of activity sampling are reported, and the advantages of sampling are stated without qualification as:

1. While it does consume the major part of the observer's day, it does not tie him up a full eight hours.... clerical time is much less than required by time study.
2. When studying jobs involving more than one worker, where two or more time study men would be necessary to get a good picture of the job, one observer can record readings for all the men in the crew.
3. It gives a more accurate picture of the job because it cannot be "padded." It should be noted that some of the service jobs can be very misleading... as to what is and what is not necessary work.
4. It relieves the man on the job of the often distasteful ordeal of "being timed" for long periods (61, pp. 75-76).

Gaylord found that he was unable to record ratings of performance while making observations, and he advocated sampling techniques for establishing standards for jobs not "lending themselves to the elemental method of analysis" (61, p. 76). The advantages listed above were not qualified, and the basic differences in the measurement objectives of elemental time study and activity sampling are not noted.

At the 1953 Industrial Engineering Institute held at the University of California, Gustat (63) reported the use of activity sampling for

determining standards for indirect labor consisting of a warehouse material handling group. Satisfactory results are reported, and alignment charts or nomographs are advocated for sample size calculations. This paper did not present new theory or experimental evidence, and the interdependence of activity was not treated. In May of 1953, Tippet (150) described his early activity sampling studies in Time and Motion Study, and in this presentation he brought out the importance of defining the elements or activity classification in clear detail. Other difficulties encountered include those related to locating the operator in cases where the operator moved about to some extent. He stated:

Precaution against similarities between the routine of the operatives and observers must be taken for all observations. Ideally, of course, the snap readings would be taken at random instants, but since anything other than a fairly regular routine is impracticable, it is obviously important that any periodicity in the observations should not coincide with periodicity in the events recorded. This requirements presents no difficulties" (150, p. 13).

In a section describing the experience of the Shirley Institute with "snap observations," Tippet states:

Theoretically it is possible to rate snap readings and my colleagues have ideas for developing methods, but so far rating has not been done. If operatives sometimes fumble over a job, or make a false start, the snap reading time includes these effects whereas the time study man will usually ignore "muffed" jobs and will time only "normal" ones. All operatives are represented in the snap reading average: experienced and inexperienced, skilled and unskilled; the time study engineer will usually not time "abnormal" operative. I do not say which part of element time is preferable or in what circumstances one is preferable to the other; I merely point out the differences (150, pp. 14-15).

The binomial model is employed and the usual expression for the standard error of the estimate is given. Tippet pointed out that the estimate is for the true fraction, "p for the machines and operatives and for the period of time over which the observations are spread" (150, p. 16).

Control charts are advocated for ascertaining stability.

Smith (144) presented results of an activity sampling study conducted at the Lincoln Electric Company to determine the utilization of ten overhead cranes. Activity was divided into five categories, four work and one idle. This was the first published study related to material handling operations but otherwise it was an insignificant contribution. In June of 1953, Barnes, Buffa, and Gary (21) reported results from a performance sampling study. The objective was an investigation of the distribution of performance ratings when observations are made over a long period of time. A total of 8,000 observations were made, and it was found that the distribution was non-normal. The Pearson, Chi-Square test was employed for the test of conformity. This was the first paper which dealt with performance sampling. Barnes presently defines work sampling as a combination of activity sampling and performance sampling.

In July, 1953, Mac Niece (90) presented a testimonial for activity sampling which was insignificant, and in August, McAllister (103) presented a paper advocating the use of random number tables for selecting observation instants. The sampling scheme suggested by McAllister is a simple random scheme with no stratification by day. Instants are in one minute increments which would give rise to a hypergeometric model rather than the binomial model assumed by MacAllister. Davidson (46) presented a paper before the fall ASME meeting in 1953. He reviewed the historical development of activity sampling noting the logical errors contained in much of the work of Brisley, Correll, Barnes and Morrow. Davidson recognized the inference problem by noting the difference between statistical estimation and forecasting. Estimation is shown to be the drawing of

inference about the time interval over which sampling is accomplished, and extrapolation into the future is defined as forecasting. Errors in the previous literature relative to the logical framework regarding replication and inference were presented. This is a logical argument, and no experimental results are presented.

In September, 1953, Eastman (53) reported the use of activity sampling for determining the fraction of idle time for a group of 99 clerical workers. This was the first reported application of sampling methods for the study of office work. Nadler (115) authored a paper, published in October, 1953, in which activity sampling is advocated for determining the time utilization of foremen and assistant foremen. He seemed to be unaware of the works of McAllister and Nelson, Davidson, and Tippet, and he states:

Rule 3. The minimum time interval between observations should be equal to the maximum time for the longest job performed by the individual. This maximum time, of course, is an average value (115, p. 91).

Matthis (101), in a short note, advocated activity sampling for office worker utilization studies, and in December, 1953, Business Week magazine presented an editorial in which activity sampling was described; however, these articles were not research papers, and they do not represent a substantial contribution to the existing literature. In 1954, numerous studies were reported. Owens (122) authored a paper in which he advocated a certain type of systematic or fixed interval sampling. This paper represents a departure from the theoretical model used by earlier authors. Owens treats the problem of systematic sampling where a certain class of activity called "interrupter activity" is studied. Interrupter activities

are defined as those activities having duration which is independent of the overall job cycle and closely related to the specific interrupter activity under study.

Gaylord and Gillespie (62) described how they used results from activity sampling study to establish incentive standards at the Aluminum Company of America. They compared results obtained from sampling with those obtained from production study, and "reasonable" agreement was displayed. The sampling design is stratified random sampling with days serving as the basis for stratification. The variance calculation was based on the binomial model which will overstate variability in this case.

At the Industrial Engineering Institute in February of 1954, O'Neill and Carrabino (120) presented the results of their study of a cargo-handling system. Their objective was:

To devise equations for the flow of material through a particular ideal cargo-handling system. These equations are to be used for the prediction of the time required to load or unload cargo or for the development of optimum methods and designs (120, p. 81).

Activity sampling for determining element times for cycles was employed, and a cluster sampling design was utilized in which the activities within a cluster are interdependent. The binomial model is erroneously assumed in determining the precision of the estimates in that the activities are not independent. At the same institute, Barnes, Buffa, and Demangate (24) presented some non-parametric statistical methods for the calculation of sample size in performance sampling, since their earlier work had shown non-normality of the distribution of performance indices. From the Tchebycheff inequality,

$$n = \frac{\sigma^2}{d^2 \alpha}$$

where n is sample size, d is the desired half-interval width, σ^2 is the variance of the performance level population, and α is the level of significance. Since σ^2 is usually unknown, an iterative procedure is suggested for determining the sufficiency of the sample. The procedure for sample size calculation involves the use of the upper bound on the confidence interval on σ^2 which is established from a preliminary sample. This procedure does not stand up under criticism. The problem of statistical instability is not treated.

The Management Review (99) featured an editorial in which activity sampling methods are advocated, and the results of a study in Hudson's Department Store of Detroit, Michigan are presented. Rowe (137) presented a paper in 1954 in which he described activity sampling and reviewed the historical development of the technique. Stratified random sampling is suggested with the inappropriate binomial variance model employed.

In March of 1954, Kinniburgh and McTaggart (85) presented the results of several activity sampling studies in the building industry. There is no treatment of variation. Groups of individuals with interdependent activities were studied by observing the activity of each member of the group at a given instant. This sampling scheme, known as cluster sampling, was employed for both the fixed interval and random interval studies reported. Continuous time study was used to check results and the authors show good agreement between sampling results and results from continuous stop-watch studies for both designs. Allderige (11) authored a short paper in which he advocated control chart utilizations with nomographs used for calculating sample size and control limits, and he presented nomographs with instructions for application.

In May 1954, The Journal of Industrial Engineering published the first of a two article series authored by Malcolm and Sammett (98, p. 141). The first article outlines the applications for which sampling methods had been employed to the date of publication, and three aspects of the problem of statistical tests and analysis are noted as follows:

1. Tests of reliability of the proportions derived from a given study (where a given study is defined as the procurement of several hundred observations, closely spaced over a continuous relatively short time period).
2. Tests of homogeneity of results of a series of studies, (the repetition of "given studies" over time or in relation to known variations in the work situation), of a given job over a time period...; or analysis of variations observed in such studies.
3. Evaluation of the significance of work sampling results as "forecasts." A correct appreciation of the applicability and limitations of statistical tests with respect to work sampling results and the design of observation procedures rests on the careful distinction between these categories (98, p. 6).

The authors state that for the first category, tests of reliability of the proportions derived from the particular given study may be made on the basis of approximation to the binomial theory (98, p. 6); and in the second category they state:

Statements regarding the reliability of estimates derived from the series of studies, however, require further analysis. This may involve analysis of variance, regression analysis, or other procedures to test the homogeneity of the group of studies or to measure variation in the results in relation to time, or with respect to some measured variation in the work situation (98, p. 7).

The authors advocated recording long delays as many times as they are observed.

The second article deals primarily with the interrelationship between sample size, half interval width and the parameter estimate; however, there is an excellent clarification of the conceptual "population."

Such populations may take numerous forms, including the following:

1. Events in a single, continuous time period in which variations in the situation being observed are random.
2. Events in a single, continuous time period in which trend is, or may be, present.
3. Events in discontinuous time periods - for example, in situations that occur seasonally, in this case, the events occurring in a particular season might constitute a subpopulation or strata, comparable to types 1 or 2.
4. Events in which a time sequence is not significant. For example, in studies of a given job performed under known variations in the work situation, the population may consist of certain values - for example, production unit times - associated with all possible variations in the work situation. The events occurring during the observation period for a particular variation in work situation studies may constitute a subpopulation such as described in types 1, 2, and 3.
5. Events related to a particular job performed in several different plants. In this case, the population may consist of certain values related to all jobs of this particular type in all plants. Again, subpopulations may be defined which consist of all events occurring with respect to this job in a particular time period in a given plant...

In the simplest case, type 1, the probability of occurrence of a given event is constant throughout the observation period, and tests based on the binomial law are appropriate. When trend is, or may be, present as in type 2, different probabilities may be associated with particular sectors of the total observation period. Even so, if the population is regarded as a continuum containing a very large number of individual events - each of which has an equal opportunity of being observed - work-sampling data should yield an unbiased estimate of the average performance for the period, and tests of reliability based on an approximation to the binomial law should be applicable to the estimated average value.

The presence of trend during a given time period can be ascertained by making a series of separate studies, distributed over the time period. If appropriate tests indicate random variations between the results of the separate studies, the data may be pooled. Depending on the circumstances, this may involve pooling all the observations from all the separate studies or pooling the results - for example, delay proportions or unit production times - of the separate studies. When time trends are present, regression analysis can be applied to the results of the individual work-sampling studies. This would provide an estimate of the amount of the time trend and a basis for tests of significance of this estimate.

In such cases, a series of work-sampling studies may be made, with each study related to a known - or selected - variation in the work situation. As in the preceding example, non-parametric tests, analysis of variance, or graphic analysis may be used to test the homogeneity of results from the separate studies. Depending on the indications such tests, the separate studies may be pooled or the nature of the relationship to the known variation in work situation may be estimated (141, p. 9).

Statistical tests are then divided into two categories as:

1. Simple tests, based on the binomial law and applied to data from individual work-sampling studies, or similar tests applied to the pooled observations from several studies if this procedure is indicated by a preliminary analysis, and
2. Tests of significance applied to the results of analysis of data in which the variation between separate studies is not random (141, p. 9).

These papers are among the few thoughtful presentations made.

In June 1954, Abdellah and Levine (1) authored an article in which they reviewed techniques which had been advocated for determining the utilization of nursing personnel. Three methods making use of continuous observation and activity sampling were reviewed. The problems of variation were not treated; and in the section presenting results obtained in several Michigan hospitals, the experiment was inadequately described. It would be difficult to replicate this work. Evidently, systematic sampling was employed as the authors state: "All unit personnel were observed every fifteen minutes over a twelve hour period from 7 A.M. to 7 P.M." (1, p. 14). The sampling was stratified by day, and there was no indication of the method used in selecting the time for the first observation each day.

Jones (80) authored a short note, published in July, 1954, in which sampling techniques are recommended for dealing with problems related to machine interference. Allen (15) described an application of activity

sampling to the activity of foremen at the Alcoa Company. Stewart (146) described activity sampling as a method of determining the time utilization of office employees, and Mahaffey (97) presented the sampling method used by the Eastman Kodak Company for determining standards for indirect labor. None of these articles was a major contribution to the literature.

Wright (163) authored a book in which she presented results of an extensive study of patient care in hospitals of the Detroit area. She states:

Work sampling, otherwise known as "statistical ratio delay study," also appeared to be well adapted to this situation, and it turned out that work sampling techniques showed up as a good method of demonstrating the principles of industrial engineering, in addition to providing an ideal tool to be used in over-all surveys of non-repetitive jobs (163, p. 87).

The results of a study are presented, and although there is no detailed description of experimental methods, the following statement is pertinent:

It was now decided that our purposes would best be served if we could have observations taken once every thirty minutes around the clock. These rounds would not be made at regular intervals but merely at some time during the interval (163, p. 116).

In January, 1955, Rowe (138) authored an article in which he illustrated the difficulties arising from meeting sample size requirements when the requirement for half-interval width is stated as a fraction of the parameter being estimated and the parameter, p , is small. It is pointed out that if a relative accuracy of 5 per cent is required, the fraction $p = 0.01$, and a confidence level of 95 per cent is used; then 158,000 observations would be required (138, p. 61). Rowe suggested expressing half interval width requirements in absolute terms, and he concluded that this "provides a more reasonable study commensurate with the accuracy of the other components of time standards" (138, p. 64). In the

opinion of the writer, this conclusion seems to be justified. Jaske (83) authored an article in which he gave a testimonial for activity sampling and described its use at the A. E. Staley Manufacturing Company of Decatur, Illinois, and Maguire (119) described an application activity sampling for determining the fraction of time spent in unsafe activity at the Monsanto Chemical Company.

Chemical Week (37) published an editorial describing the use of activity sampling for determining the utilization of maintenance employees in the chemical industry, and Gustat (63) presented a paper before management division of the American Society of Mechanical Engineers in which he described a sampling study of warehouse handling group. Mark sense cards were employed for recording the date, operation, rating, and results.

At the 1955 conference of the American Institute of Industrial Engineers, Brisley (29) presented a paper praising the activity sampling technique. The paper was not original and Brisley states: "much of what I have to present to you is not new to you for I stole it from someone else" (29, p. 168).

Heiland and Richardson (72) coauthored an article describing "an example of the use of the technique of work sampling as a fact finding device which can help in effective office management" (72, p. 1157). The authors conclude that the technique, if properly applied, should serve as a basis for sound decisions on staff and equipment needs and help identify work which is erratic.

Macomber (91) advocated activity sampling for time-utilization studies, and he presented typical results obtained by the A. T. Kearney Company, in the study of maintenance activities.

Davis (49) authored an article comparing three methods of selecting times for observation in activity sampling. This article presents a rigorous mathematical treatment, and it constitutes a real addition to the basic knowledge of activity sampling. It was reviewed in detail in Chapter II of this thesis.

Factory magazine presented an editorial (54) in August, 1955, which contained the view that "Uniform allowances tacked on the leveled stopwatch time are unfair" (54, p. 112). In this article the editors advocate the application of activity sampling for determining the delay allowances. No data are presented, and the article is insignificant.

White (157) authored an article published in October of 1955 in which he described activity sampling as a method which "enables management to get facts without watching everything and everybody all the time" (157, p. 238). The paper described the experience of White at the John Deere Waterloo Tractor works.

Niebel (117) described activity sampling, utilizing a binomial model, as a method for determining delay allowances. He argued that management is in a difficult position with labor relations when elemental time study is followed by some rough estimate of delay allowances (117, p. 115). The article is insignificant in so far as the basic knowledge is concerned, and Niebel's text (118) published in 1955, inadequately describes activity sampling.

Barnes and Andrews (19) reviewed accomplishments in performance sampling. They compared normal time values calculated by stopwatch methods to those calculated by utilizing activity sampling and performance sampling. Experimental validation studies are presented, and the authors conclude:

Our studies seem to indicate that work sampling will give time standards for repetitive standardized manual tasks which are substantially the same as standards obtained by time study (19, p. 8).

In view of the experimental evidence presented, this conclusion seems justified; although no precise definition of "substantially the same" is given. Sample size problems associated with the performance sampling are treated in the same manner as in Barnes' earlier work on this topic (21, 23, 24).

Addoms and Long (6) authored a paper in which an experiment is reported; however, there is a clear lack of understanding of the philosophy of statistics. After stating half interval and confidence requirements and calculating the sample size required based on a binomial model and an estimated smallest fraction of time spent in any one activity, the authors glibly change their requirements in order to reduce sample size. In their presentation concerning the choice of moments for observation, it is decided to allow one minute for recording the observation of activity and a rating for performance level, and the sampling design was simple random sampling (6, pp. 30-31). An alternate method of selecting observation instants via punched cards is presented; however, in this case this procedure does not seem to be economically justified.

In his text, Mundel (112) advocated activity sampling for determining allowances to be used in connection with elemental time study data for establishing work standards. A binomial model and simple random sampling are advocated; however, this is generally an inadequate presentation as was that in Nadler's text (116). Both authors seem to lack an understanding of random variation as well as a knowledge of the previous literature dealing with time-to-time variation in activity sampling.

The Rand Corporation sponsored the publication (130) of a table of

1,000,000 random digits in 1955. This table has been used to a large extent by practicing industrial engineers working with the determination of sampling procedures. Hatton (70) authored the first British paper dealing with performance sampling; however, it is not a significant contribution to the literature. Cote and Scott (44) presented an outstanding article comparing all day time study and activity sampling and this paper was reviewed in detail in Chapter II of this thesis.

In February, 1956, Anderson (16) presented a paper at the Middle Atlantic Conference of the American Society for Quality Control in which several applications of activity sampling are noted. This paper was not a report of research activity. Case studies reported by Voris (153) and Agett and Walker (7) were of no significance from a research standpoint.

Barnes and Andrews (19) presented a paper before the Eighth Industrial Engineering Institute in February, 1956; however, the information presented had been published in the earlier work of Barnes et al., and this paper adds nothing to the earlier work. In May, 1956, Barnes (20) read a similar paper before the annual Conference of the American Institute of Industrial Engineers in Washington, D. C. Neither of these presentations was a significant contribution in view of the fact that the information presented here had been published earlier.

In June, 1956, Heller (74) gave a paper in which he advocated activity sampling and presented some illustrative applications taken from the files of the Kingston, North Carolina, Dupont Plant. The paper does not represent a significant contribution.

Moder and Halladay (107) considered the case where teams or crews are at work on long cycle tasks. The case studied was that of the activity

of individuals within crews being correlated and activity between crews being independent. This paper was reviewed in Chapter II of this thesis.

Allderige (12) authored an article published in September, 1956, in which he advocated applying some of the "new tools" of operations research to the "old problems" of industrial engineering. The particular tool illustrated was queueing theory, and the application was that of determining the number of men for a maintenance crew. Activity sampling was employed as a first stage to the problem solution. McNaughton (104) authored a short note, published in October, 1956, in which he advocated the use of activity sampling for measuring the time utilization of executives, and Paul (124) authored a similar note published a month later in which he advocated activity sampling for office worker time utilization studies. Both of these notes indicate a lack of familiarity with random variation.

In 1956, the text, Work Sampling (22), authored by Barnes, was published. This was the first book dealing exclusively with the subject of activity sampling. No new material is presented and for the most part the presentation is made up of edited reprints from earlier papers by a number of authors. As was the case with the papers presented by Barnes, the language usage and low level technical content provide material easily understood by industrial foremen having little or no education beyond the secondary school level.

Dumene (52) presented a paper in February, 1957 in which he advocated the use of activity sampling for studying the activity of maintenance and construction workers. No experimental evidence was presented, and no new theory was developed, and the paper was an insignificant contribution.

Williams (159) authored an informative article in which he summarized

the techniques of work study* with particular emphasis on activity sampling and its position relative to the other techniques. There is nothing original in the way of theory or experimental evidence presented in this article. Conway (42) authored an informative article somewhat similar to the previous article reviewed. This paper has been reviewed in Chapter II of this thesis. Allderige (13) presented a paper in which he explained the role of the mathematical model relative to measurement. His presentation reported the use of two models in work measurement, one for continuous activity study, and one for activity sampling. The nature of the variation present is elaborated; however, no experimental evidence is presented. Emphasis is placed on the control system relative to work measurement. Barnes (18) presented a paper before the twelfth management engineering conference in which he reviewed the contents of his text (25) published the year before, and Underhill (151) reported an application of activity sampling for determining the time utilization of foremen. Underhill refers to the path taken by the observer through the work area as a random path; however, there is no description of how such random paths are generated, and there is no treatment of variation.

French (59) authored an article describing an application of activity sampling in the food processing industry, and Barlament (17) authored a note in which he described the use of random interval photographic device for filming observations. No details are given as the note is quite brief. This work was published by Textile World in September, 1957, and it was followed by a companion note the following month which was authored by Brownstein (34). This latter note presents some data which were obtained

*

Defined by Williams to be the study of worker or machine behavior.

from a study of fifty sewing machine operators. No new analysis methods or theoretical developments were presented.

Basu and Dastidar (26) authored an article describing activity sampling for application in the factories and plants of India. No new concepts are presented, and the article is presented on an elementary level; however, it is informative in nature and presents enough examples to be easily understood by persons with relatively little technical training.

Heiland and Richardson (73) presented a text which was published in 1957 by the McGraw-Hill Book Company. This book does not give a technical treatment of the subject matter, and the authors have attempted to make the material understandable for individuals with little formal education. Several of the articles published prior to the time of this publication are reviewed in the text.

Morrow (110) presented a second edition of the book, Motion Economy and Work Measurement, which was published in 1957. Two chapters entitled "Ratio-Delay Study" (110, p. 273-296) and "Work Sampling" (110, p. 297-319) are devoted to a description of simple random sampling and the binomial model applied to activity analysis. No new concepts are presented; however, a number of interesting case studies are included.

Breeze (28) authored a short note advocating activity sampling as an appropriate technique for the setting of work standards. The presentation does not contribute to the existing knowledge of this subject.

Rosander, Guterman, and McKeon (136) authored an excellent article in which they presented the methods employed by the Internal Revenue Service in conducting activity sampling studies, and Heller (74) read a paper at

the ninth annual Industrial Engineering Conference at the University of West Virginia in which he mentioned many of the successful applications of activity sampling at DuPont and other companies.

Haines (64) presented a comparison of results obtained from fixed interval study and continuous study, and he found good agreement. Page (123) authored an article describing a methodology for utilizing key sort cards for recording activity sampling observations.

In February, 1959, Peterson (125) presented a paper at the Eleventh Annual Industrial Engineering Institute at the University of California in which the PACE* program of the Northrop Aircraft Company was described. The formula for computing the PACE index as a function of the four input measurements was presented, and sampling was recommended for obtaining the input measurements 2, 3, and 4 defined in the footnote. The paper does not give complete details, so that it is difficult to evaluate the method advocated. No experimental evidence is presented.

In April, 1959, Reul and Richardson (131) advocated the application of activity sampling for isolating problem areas in material handling operations. Illustrative results are presented; however, this is not a rigorous presentation, and variation is not considered.

Klein (86) authored an article in which he treated the problems related to the determination of sample size to satisfy an established requirement for confidence interval width when the parameter, p , is unknown. A binomial model is assumed, and Stein's two stage procedure is employed.

* A program designed to increase group work effectiveness by rating (1) The number of people assigned, plus or minus loans. (2) The average effort of the people who are working. (3) The number of people idle. (4) The number of people not in the assigned work area.

In July, 1959, Mindlin (106) presented an outstanding article in which he developed the procedures used by the social security system for activity sampling. This paper has been reviewed in Chapter II of this thesis. Results are presented with the seemingly justified conclusion that:

The method is statistically more valid and reliable than classical time study, much less time consuming and less costly to operate, and inspires genuine confidence in the field staff in the accuracy of the results (106, p. 295).

Wolff (162) described a method of fixed interval recording, where the observer and recorder are linked by a portable two-way radio.

Richardson (132) summarized several applications of activity sampling in an article published by Factory magazine, and Weir (155) wrote of an application of activity sampling in a nail mill. Neither of these papers is significant. Ritchey (133) authored an article, published in November, 1959, in which he used a stratified sampling scheme to study the activity of a university faculty. The paper presents a novel application of activity sampling, but it is otherwise insignificant. Bryant (35) described the detection system employed by United Airlines for locating ineffective work activity in their maintenance facility. Activity sampling was credited indirectly with an increase of 43 per cent in "productivity." Many of the terms used were ill defined and the paper is difficult to evaluate.

Recent Literature

Murray (113) authored an article, published in February, 1960 in which he elaborated on the problems associated with establishing standards for clerical workers; and he advocated the combined application of activity sampling and predetermined elemental times for the solution of these

problems. This is not an experimental report, and emphasis is on the argument for the application of sampling. Lifson (88) reviewed some of the applications and limitations of activity sampling in a short paper read at the Eleventh Annual Conference of the American Institute of Industrial Engineers.

Provost (128) conducted research comparing "the use of memomotion film for analyzing non-work activities with three other sampling methods; a concurrent systematic interval study, a systematic sampling of the film, and a random sampling of the film" (128, p. 1). A memomotion camera having a film speed of 25 frames per minute was used to record the activities of eight men fabricating small assemblies, and a concurrent fixed interval study, employing a two-minute interval, was conducted by an observer in the production area. Two systematic sampling studies were made from the film using a frame counter and every 50th frame starting with frame fifty, and with frame 25. Two independent simple random samples were selected from film frames, each consisting of 500 frames. Results were tested for significant differences for systematic vs. random frame sampling and no significant differences were found. A non-parametric test was conducted to ascertain which of the methods most nearly approached the true parameter obtained from a total frame count. The statistical analysis seems weak in that inter-individual variability is not separated from the main effect (method of analysis) variability.

Rosander (135) authored an excellent article, published in June, 1960, in which he compares cluster sampling and simple random sampling from an activity array defined as $[A_{ij}]$ where $A_{ij} = 1$, or 0 according to whether or not the element does or does not belong to the class of interest.

$[A_{ij}]$ is $N \times M$ matrix where the N rows represent employees and the M columns represent minutes. This presentation does not report experimental evidence, as it is a theoretical development. Cost arrays $[C_{ij}]$ corresponding to $[A_{ij}]$ are developed, and the economics of stratification are presented. The presentation is significant in that it clearly indicates the conceptual difference between cluster sampling and simple random sampling.

Schmid (143) authored a note describing the "New Schmid Work Measurement Sampling System" which is a sales oriented pitch claiming a 90 per cent reduction in work measurement costs.

Davidson, Hines, and Newberry (47) developed a formula for the variance of the estimate calculated from a systematic sample. This paper was reviewed in Chapter II of this thesis.

Davidson (48) enumerated eleven fallacies found in the literature regarding activity sampling as:

1. Work sampling works because the law of probability works.
2. Work sampling permits delay allowances to be set with a specified probability that the error will not exceed a predetermined amount.
3. Work sampling is based on the percentage number of delays, not upon their duration.
4. When developing a schedule of observations by use of a table of random numbers, or similar means, one should ignore cases where additional observations are to be taken of the same operator at the same time that an observation is already scheduled.
5. Observations would be separated by a sufficient period of time so that the same delay is not counted twice.
6. As a general rule the number of observations should be such as to provide 95% assurance (2 sigma confidence interval) that the estimated percentage of time devoted to any particular class of activity is within $\pm 10\%$ of the true percentage.

7. If n rounds of observations are taken on a crew composed of k men and an observation is made on each of the k men in each of the n rounds, then the size of the sample to be used in calculating the error of estimate is nk .
8. In nearly all cases the occurrence of delays follows the binomial probability distribution.
9. For a given number of observations, random sampling will have a smaller error variance than any other kind of sampling.
10. Randomization of the times of observation eliminates biases in the estimates of per cent time spent in various activities.
11. The main reason for randomizing observation times is that it is necessary to satisfy the requirements of statistical theory used in calculating confidence intervals (48, p. 367).

In September, 1960, Isherwood (77) presented a method used for activity sampling in a small plant nursery business on the west coast. The study employed a fixed interval sample with the binomial variance relationship for explaining the precision; however, no assumptions or limitations are given, and there is no consideration given to the possibility of underlying periodicity bringing about an increase in the variance of the estimate of the fraction of time in the various categories considered.

Millikin (105) stated that the Aircraft Division of the Twin Coach Company of Buffalo, New York utilized activity sampling and "increased effective use of productive time by 16%" (105, p. 148). Alenik and Koepp (8) stated that the Chance Vought Aircraft Company has successively employed activity sampling; however, their article is little more than a testimonial, and it is insignificant. Halsey (67) developed a model for activity sampling which represented an addition to the existing theory and a means for reducing the variability of the estimate of the fraction of time that the activity is in a given class or state. The method is that of ratio estimation employed prior to this time by survey samplers.

The primary argument for this method is that it is the correct model when observation tours are conducted and group sizes vary. The following conclusions seem justified.

1. The method allows for more flexibility and simplicity in the operational aspects of work sampling.
2. It is claimed to give greater precision for a given cost.
3. It provides a mathematical model and technique which are consistent and eliminates the necessity for deviating from a model because of practical considerations.
4. It eliminates the need to find specific people in the population under study.
5. It eliminates the problem of missing people in a mobile work force engaged in activities that require travel, changing work locations and variable tasks.
6. It is based on a somewhat more complicated model, but an extremely simple method for estimating sampling variation using the range is given (67, p. 508).

The third and last of the books published on the subject of activity sampling was authored by Hansen (68) and published in 1960. The book is written for individuals having no technical training, and compromises are made in the technical completeness and correctness in order to maintain an elementary presentation. The material presented does not represent an addition to the existing knowledge.

In January, 1961, Drui (51) authored an article in which he reported the results of a research study of the effects of pace on activity sampling results. Four jobs having common traits were selected in four different industries. The criteria for selecting the jobs were:

1. The job is generally repetitive in nature.
2. Five or more operators are assigned to the same operation.
3. The performance of the job does not require a crew; that is, each of the operators assigned to the job is performing the same work at different work stations.

4. Each operator is fully trained according to the concept of full training for the company.
5. The work of the job is manual and fully controlled by the operator.
6. The operators working on the job are paid on an incentive basis if possible.
7. The work is continuous enough to permit observations to be made over several weeks.
8. Incoming material quality is reasonably constant (51, p. 29).

There were 7,610 observations and sixteen operators on the first job, 2,455 observations and seven operators on the second job, 1,783 observations and five operators on the third job, and 1,673 observations and eight operators on the fourth job. In all cases, method was not standardized. Operators were rated at the time of each observation. The average operator pace had high correlation with operator output, and the author states that "The correlation between pace and work sampling results showed the highest significance for the elements not highly repetitive and in which the operators obtained materials whenever they needed them" (51, p. 30). The author does not state the level of correlation, and the conclusions are very general in nature. The experimental evidence seems to be insufficient to support them.

Leighty (89) presented the method used by the Boeing Airplane Company for performance and activity sampling of their indirect work force including managers and engineers. No details are given, and the paper is not a contribution of significance. In March 1961, Connor (41) reported conclusions from his research in the application of sampling to estimate the requirements for nursing time at the Johns Hopkins Hospital. In addition to time utilization studies, he studied fluctuations in the ward census,

nursing hours scheduled, and patient care required. The description of the experiment is not complete making evaluation extremely difficult. Connor concludes that "as the direct patient care increases, the time spent in direct patient care increases and the time in non-productive activity decreases" (41, p. 107). The regression technique employed by Connor was treated in a casual manner, and yet it is quite possible that no real conclusions can be accompanied by a probability statement due to the fact that the basic assumptions underlying the application of a linear regression model are not necessarily satisfied in this case.

Kruse (87) authored an insignificant note in which he praised activity sampling as an efficient way of obtaining good "rough estimates" with little sampling effort, and Wolf and Riordan (161) presented the results obtained from an extensive activity study conducted by the Bureau of Ships of the United States Navy Department. Activity codification was detailed and punch cards were employed for summarizing the data. No new methods or theoretical development were presented. Allderige (14) criticized this study on the basis that the poor design of the experiment was responsible for the large observation requirement, and he suggests that "a large part of the measurement dollars could have been saved with a small part traded for study design dollars". (14, p. 180).

Wilding-White (158) authored an article in which detailed methods for processing activity sampling data are presented. An experiment was conducted, and the author found that the observer cannot punch cards while making observations as effectively as marking a data sheet. Timmins (147) reported the application of random interval and fixed interval sampling in the British steel industry. In the random interval study,

the objective was to obtain a breakdown of the activity of casters. The random interval study employed simple random sampling. Timmins states:

A set interval system of sampling is often more convenient, particularly as a large number of the operations in foundaries for which the technique is applicable are, in fact, non-repetitive. With practice, observations can be made quite comfortably at minute or even half-minute intervals and this still leaves sufficient time to record other events, particularly if the observer has a good knowledge of the technical aspects of the procedure under examination (147, p. 133).

The fixed interval sampling procedure employed a one-minute interval, and the activity of an overhead crane was the object of the sampling.

Wheeler (156) authored an article advocating the use of time lapse photography in connection with activity sampling. This article is not technical in nature, and no details, methods, or assumptions are given. In January 1962, Van Eman (152), a maintenance engineer for Union Carbide Chemical Company, advocated activity sampling as an expedient method of activity analysis, and Moskowitz (111) suggested that performance sampling with cycle rating is superior to element rating over a continuous time period. The "tests" are vaguely described, and it is impossible to determine whether or not the author's conclusions are justified.

A P P E N D I X B

Table 12. Simulation of a Process Where Span Lengths Have Gamma Densities

```

2 COMMENT STOCHASTIC PROCESS SIMULATION GAMMA DIST 0.1 PROCESS HINIC 5
2 INTEGER I,J,K,M,N,RN,C,PPN,S,RP,R,TI,KFFD ,X 5
2 ARRAY( A(1000),AP(1000) ,X(1600) ) 5
2 INPUT DATA(LO,L,RP,R) 5
2 OUTPUT RESULT( FOR TI=(1,1,960) $ X(TI) ) 5
2 FORMAT COPY(12(454,R),10(2,40)) 5
2 OUTPUT SAVE( RM ) 5
2 FORMAT COPY(12(110,40)) 5
2 RN= 4000664853 5
2 C=1220703123 5
2 READ(55,DATA1) 5
2 I=1 5
2 P1.. FOR M=(1,1,900) $ (A(M)=0.05AP(M)+0.0) 5
2 M=C C(M)=0.0 5
2 P4.. M=M+1 $ T=0.0 5
2 FOR C=(1,1,RP) 5
2 BEGIN RN=C,RP 5
2 T= T+(ARC( LOG(10.000000001), (RN) ) ) / (1/P) ) END 5
2 ADIM=T $ SUM=SUM+ADIM $ IF SUM <= 480.0500 D5 $ GO P4 5
2 D5.. T=0.0 5
2 FOR C=(1,1,0) 5
2 BEGIN RN=C,RP 5
2 T=T+(ARC( LOG(10.000000001), (RN) ) ) / (1/P) ) END 5
2 A(M)=T 5
2 SUM=SUM+A(M) 5
2 IF SUM <= 480.0 5
2 GO P4 5
2 D6.. KFFD=M $ C=0.0 5
2 FOR M=(1,1,KFFD) 5
2 CA=CA+A(M) 5
2 WD=(CA)/(SUM) 5
2 PN=C,PN 5
2 WRN=(0.000000001),PN 5
2 IF WDN <= WD 5
2 GO P8 5
2 GO P0 5
2 P8.. SC=0.0 $ TI=1 $ FOR M=(1,1,KFFD) 5
2 BEGIN SC=SC+A(M) $ UNTIL (TI),(0.5) GTR SC 5
2 BEGIN X(TI)=1 5
2 TI=TI+1 END 5
2 X(TI)=0 $ SC=SC+A(M) 5
2 UNTIL (TI),(0.5) GTR SC 5
2 BEGIN X(TI)=0 5
2 TI=TI+1 END 5
2 X(TI)=1 END 5
2 GO P10 5
2 P10.. SC=0.0 $ TI=1 $ FOR M=(1,1,KFFD) 5
2 BEGIN SC=SC+A(M) $ UNTIL (TI),(0.5) GTR SC 5
2 BEGIN X(TI)=0 5
2 TI=TI+1 END 5
2 X(TI)=1 $ SC=SC+A(M) 5
2 UNTIL (TI),(0.5) GTR SC 5
2 BEGIN X(TI)=1 5
2 TI=TI+1 END 5
2 X(TI)=0 END 5
2 GO P10 5
2 P10.. WRITE(8,RESULT,COPY) 5
2 I=I+1 5
2 IF I <= 5) 5
2 GO P8 5
2 WRITE(8,SAVE,COPSAVE) 5
2 STOP 5
2 FINISH 5

```


Table 15. Simulation of a Process Where the Span Lengths Have Exponential Densities

```

2                                     COMMENT
2 THIS PROGRAM IS FOR THE SIMULATION OF A TWO-VALUED STOCHASTIC PROC.
2 X(T)=0,1 WHERE THE LENGTHS OF TIME OVER WHICH X(T)=0 HAVE AN EXPON.
2 DIST. AND THE LENGTHS OVER WHICH X(T)=1 HAVE AN EXPONENTIAL DIST.
2 THE MEAN SPAN LENGTHS ARE 10.0 AND 10.0 RESPECTIVELY
2 INTEGER I,J,K,M,N, C,PRN,C,RP,R,TI,KFFP,X,FRN1,FRN2,FRN3
2 ARRAY A(1000),AP(1000),RN(10),X(1600)
2 INPUT DATA(LP,L,RP,R)
2 INPUT DATA2(FOR J=(1,1,10)RN(J)
2 OUTPUT RESULT(FOR TI=(1,1,960) X(TI))
2 FORMAT COPY(2(1,64,R),30(2,40))
2 READ($DATA1)
2 I=1
2 P3.. FOR M=(1,1,999) $ (A(M)=0,AP(M)=0.01
2 M=0 C(M)=0
2 P3A.. READ($DATA2) K=1
2 P4.. M=M+1 ST=0
2 T=ABS(LOG(RN(K)))/L
2 AP(M)=T C(M)=C(M)+AP(M) IF C(M) LSS 480.05 GO P5 T GO P6
2 P5.. T=0.0
2 K=K+1 T=ABS(LOG(RN(K)))/L
2 A(M)=T
2 C(M)=C(M)+A(M)
2 IF (C(M) LSS 480.0) AND (K LSS 10)
2 ( K=K+1 $ GO P4 )
2 IF (C(M) LSS 480.0) AND (K FOL 10)
2 GO P3A
2 P6.. KFFP = M*WD=0.0
2 FOR J=(1,1,KFFP)
2 WP= WP+ AP(J) +A(J)
2 WP= WP/480.0
2 PRN1=RN(1),100000$ PRN2=RN(5),100000$ PRN3=RN(7),100000
2 FRN1=FIX(PRN1) FRN2=FIX(PRN2) FRN3=FIX(PRN3)
2 PRN = FRN1,FRN2,FRN3
2 VRN=(0.0000000001),PRN
2 IF VRN LSS WP
2 GO P3
2 GO P3
2 P7.. SS=0.0 $ TI=1 $ FOR M=(1,1,KFFP)
2 BEGIN SS=SS+A(M) $ UNTIL (TI),(0.5) GTR SS
2 BEGIN X(TI)=1
2 TI=TI+1 END
2 X(TI)=0 $ SS=SS+A(M)
2 UNTIL (TI),(0.5) GTR SS
2 BEGIN X(TI)=0
2 TI=TI+1 END
2 X(TI)=1 END
2 GO P10
2 P8.. SS=0.0 $ TI=1 $ FOR M=(1,1,KFFP)
2 BEGIN SS=SS+A(M) $ UNTIL (TI),(0.5) GTR SS
2 BEGIN X(TI)=0
2 TI=TI+1 END
2 X(TI)=1 $ SS=SS+A(M)
2 UNTIL (TI),(0.5) GTR SS
2 BEGIN X(TI)=1
2 TI=TI+1 END
2 X(TI)=0 END
2 GO P10
2 P10.. WRITE($RESULT,COPY)
2 I=I+1
2 IF I LSS 51
2 GO P3 $ STOP $ FINISH

```

Table 16. Analysis of Simulation Results

```

2 COMMENT FURTHER ANALYSIS OF SIMULATED REALIZATIONS          $
2 INTEGER X(1),J,K,L,M,N,F,C,CUM,INTV,TOP                      $
2 ARRAY(X(960),F(25),WV(25))                                  $
2 INPUT DATA1(FOR I=(1,1,18)KF(L))                          $
2 INPUT DATA2(FOR I=(1,1,960)SX(1))                          $
2 OUTPUT RF4(F(L),WV(L))                                       $
2      FORMAT (4F50.0, R10.11, R10.11, X10.4, W0)             $
2      READ(5,DATA1)                                           $
2      J=0                                                      $
2      P1=. READ(5,DATA2) $J=J+1$CUM=0$FOR I=(1,1,960)$CUM=CUM+X(I) $
2      P=(FLOAT(CUM)/960.0 $ V=P(1,0-P)                       $
2      FOR L=(1,1,18)                                          $
2      BEGIN $C=CUM$P=0.0$ INTV=F(L)                          $
2      FOR K=(1,1,1)F(L))                                     $
2      BEGIN TOP=960-F(L)+K$                                  $
2      FOR N=(K,INTV,TOP)                                     $
2      S=S+X(N)                                                $
2      P=((FLOAT(S), (FLOAT(F(L))) / 960.0 $                  $
2      S=S+((P)(1)-P) / 960.0) / FLOAT(F(L)))                  $
2      S=0 $ P=0                                              $
2      RV = (960.0 / FLOAT(F(L))) - (C / (V * F(L))) $         $
2      WV(L) = CORR(RV) $                                     $
2      WRITE(5,RF4,C4) $S=C$0.0 ENDSIF J L$C 50$FOR P)      $
2      STOP $FINISH                                           $

```

Table 17. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$,
RN = 1190000001, and $C = 7^{11}$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.81015 | 0.02507 | 31 | 0.01203 | 0.09685 |
| 3 | 0.52707 | 0.05349 | 33 | 0.01824 | 0.10324 |
| 5 | 0.33222 | 0.07477 | 35 | 0.02102 | 0.10063 |
| 7 | 0.20404 | 0.09036 | 37 | 0.01882 | 0.09746 |
| 9 | 0.11837 | 0.10057 | 39 | 0.00967 | 0.09177 |
| 11 | 0.06032 | 0.10255 | 41 | -0.00041 | 0.08396 |
| 13 | 0.02088 | 0.09849 | 43 | -0.00211 | 0.08440 |
| 15 | -0.00180 | 0.09451 | 45 | -0.00708 | 0.08979 |
| 17 | -0.01427 | 0.09742 | 47 | -0.01951 | 0.10417 |
| 19 | -0.01555 | 0.09423 | 49 | -0.02875 | 0.12126 |
| 21 | -0.01439 | 0.10037 | 51 | -0.03602 | 0.11972 |
| 23 | -0.00741 | 0.09605 | 53 | -0.03877 | 0.11497 |
| 25 | -0.00463 | 0.09373 | 55 | -0.03917 | 0.11458 |
| 27 | -0.00003 | 0.08846 | 57 | -0.03973 | 0.11667 |
| 29 | 0.00316 | 0.08670 | 59 | -0.04436 | 0.11298 |

Table 18. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$,
RN = 2559066073, and $C = 518$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.81635 | 0.02877 | 31 | 0.04123 | 0.11684 |
| 3 | 0.54560 | 0.07089 | 33 | 0.04851 | 0.11196 |
| 5 | 0.36224 | 0.08515 | 35 | 0.03026 | 0.09544 |
| 7 | 0.24008 | 0.08033 | 37 | 0.01902 | 0.09385 |
| 9 | 0.16112 | 0.08497 | 39 | 0.00301 | 0.10476 |
| 11 | 0.10455 | 0.09156 | 41 | -0.01280 | 0.11253 |
| 13 | 0.06792 | 0.11098 | 43 | -0.04159 | 0.11550 |
| 15 | 0.03667 | 0.12333 | 45 | -0.06785 | 0.10630 |
| 17 | 0.01098 | 0.11939 | 47 | -0.07167 | 0.09660 |
| 19 | -0.01718 | 0.11311 | 49 | -0.06425 | 0.08961 |
| 21 | -0.04243 | 0.10290 | 51 | -0.06105 | 0.09209 |
| 23 | -0.06338 | 0.10281 | 53 | -0.07029 | 0.10063 |
| 25 | -0.05513 | 0.10568 | 55 | -0.08619 | 0.10088 |
| 27 | -0.03843 | 0.10890 | 57 | -0.08354 | 0.10036 |
| 29 | 0.00263 | 0.11712 | 59 | -0.08029 | 0.10949 |

Table 19. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$,
RN = 4999465853, and $C = 5634765623$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.821091 | 0.023190 | 31 | -0.049257 | 0.096923 |
| 3 | 0.546641 | 0.059340 | 33 | -0.048635 | 0.097017 |
| 5 | 0.356795 | 0.087302 | 35 | -0.042811 | 0.094861 |
| 7 | 0.225773 | 0.101816 | 37 | -0.034573 | 0.102461 |
| 9 | 0.142551 | 0.104298 | 39 | -0.025905 | 0.120697 |
| 11 | 0.090013 | 0.102195 | 41 | -0.016813 | 0.130792 |
| 13 | 0.053996 | 0.095178 | 43 | -0.017737 | 0.129717 |
| 15 | 0.028615 | 0.098966 | 45 | -0.022439 | 0.120243 |
| 17 | 0.010956 | 0.101288 | 47 | -0.021478 | 0.117461 |
| 19 | -0.008623 | 0.102549 | 49 | -0.021624 | 0.116231 |
| 21 | -0.012470 | 0.098471 | 51 | -0.017731 | 0.108297 |
| 23 | -0.014444 | 0.098770 | 53 | -0.019493 | 0.105788 |
| 25 | -0.025657 | 0.090702 | 55 | -0.023485 | 0.104627 |
| 27 | -0.037161 | 0.087523 | 57 | -0.021268 | 0.106761 |
| 29 | -0.047780 | 0.090284 | 59 | -0.022077 | 0.116093 |

Table 20. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 0.1$, $r = r' = 1$,
and Random Number Input

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.815872 | 0.028177 | 31 | -0.014012 | 0.121194 |
| 3 | 0.540689 | 0.068066 | 33 | -0.013313 | 0.118879 |
| 5 | 0.357774 | 0.089310 | 35 | -0.015014 | 0.121532 |
| 7 | 0.240375 | 0.098754 | 37 | -0.011929 | 0.117909 |
| 9 | 0.160331 | 0.112111 | 39 | -0.14282 | 0.111992 |
| 11 | 0.112992 | 0.110201 | 41 | -0.008554 | 0.109829 |
| 13 | 0.088625 | 0.107777 | 43 | -0.008278 | 0.111943 |
| 15 | 0.071639 | 0.116424 | 45 | -0.009872 | 0.113429 |
| 17 | 0.064907 | 0.120843 | 47 | -0.006873 | 0.106093 |
| 19 | 0.051756 | 0.120948 | 49 | -0.013213 | 0.104435 |
| 21 | 0.040497 | 0.114800 | 51 | -0.016770 | 0.109812 |
| 23 | 0.029501 | 0.113346 | 53 | -0.018668 | 0.106895 |
| 25 | 0.015659 | 0.104208 | 55 | -0.016356 | 0.109319 |
| 27 | -0.001902 | 0.101232 | 57 | -0.023306 | 0.123751 |
| 29 | -0.014037 | 0.112438 | 59 | -0.019623 | 0.136396 |

Table 21. Values of the Correlation Function from
Simulation for $\lambda = 0.4$, $\lambda' = 0.08$,
 $r = 2$, $r' = 2$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.764745 | 0.038960 | 31 | 0.009774 | 0.081231 |
| 3 | 0.384246 | 0.083674 | 33 | 0.016978 | 0.076953 |
| 5 | 0.145093 | 0.079777 | 35 | 0.014651 | 0.088967 |
| 7 | 0.014048 | 0.066393 | 37 | 0.019734 | 0.094247 |
| 9 | -0.050868 | 0.060006 | 39 | 0.017358 | 0.082618 |
| 11 | -0.080990 | 0.058446 | 41 | 0.011886 | 0.076443 |
| 13 | -0.086378 | 0.064495 | 43 | 0.006961 | 0.079852 |
| 15 | -0.082030 | 0.059765 | 45 | -0.015009 | 0.079585 |
| 17 | -0.072482 | 0.061744 | 47 | -0.030387 | 0.085909 |
| 19 | -0.056616 | 0.065532 | 49 | -0.027772 | 0.084597 |
| 21 | -0.035331 | 0.077818 | 51 | -0.018619 | 0.084340 |
| 23 | -0.020206 | 0.078209 | 53 | -0.022568 | 0.089886 |
| 25 | -0.010764 | 0.081744 | 55 | -0.020665 | 0.091056 |
| 27 | -0.002549 | 0.084431 | 57 | -0.015501 | 0.089897 |
| 29 | 0.000314 | 0.086882 | 59 | -0.017955 | 0.083678 |

Table 22. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 1.0$,
 $r = r' = 10$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.799717 | 0.009070 | 31 | -0.179391 | 0.120444 |
| 3 | 0.397206 | 0.027239 | 33 | -0.118122 | 0.122527 |
| 5 | 0.002264 | 0.044526 | 35 | -0.035139 | 0.130897 |
| 7 | -0.323278 | 0.052914 | 37 | 0.043572 | 0.135791 |
| 9 | -0.492715 | 0.045318 | 39 | 0.087721 | 0.134854 |
| 11 | -0.450661 | 0.062880 | 41 | 0.096723 | 0.129126 |
| 13 | -0.256163 | 0.085850 | 43 | 0.078068 | 0.124191 |
| 15 | -0.010012 | 0.103488 | 45 | 0.039054 | 0.128445 |
| 17 | 0.198520 | 0.098303 | 47 | -0.011858 | 0.137512 |
| 19 | 0.297746 | 0.086558 | 49 | -0.047853 | 0.140930 |
| 21 | 0.280383 | 0.093656 | 51 | -0.065132 | 0.139748 |
| 23 | 0.169393 | 0.109297 | 53 | -0.059046 | 0.135826 |
| 25 | 0.013356 | 0.126594 | 55 | -0.043348 | 0.141772 |
| 27 | -0.112110 | 0.128959 | 57 | -0.019077 | 0.147383 |
| 29 | -0.184183 | 0.125772 | 59 | 0.011840 | 0.143153 |

Table 23. Values of the Correlation Function from
Simulation for $\lambda = 3.0$, $\lambda' = 2.0$,
 $r = 9$, $r' = 14$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.521517 | 0.018047 | 31 | 0.034373 | 0.072029 |
| 3 | -0.241315 | 0.027535 | 33 | -0.006253 | 0.083199 |
| 5 | -0.369453 | 0.029266 | 35 | -0.039314 | 0.067416 |
| 7 | -0.102676 | 0.063188 | 37 | -0.016061 | 0.075570 |
| 9 | 0.223794 | 0.054743 | 39 | 0.017593 | 0.079514 |
| 11 | 0.196426 | 0.079724 | 41 | 0.029207 | 0.070532 |
| 13 | -0.060689 | 0.063081 | 43 | -0.003117 | 0.082934 |
| 15 | -0.167402 | 0.069358 | 45 | -0.019863 | 0.066429 |
| 17 | -0.055410 | 0.062989 | 47 | -0.020450 | 0.078674 |
| 19 | 0.098502 | 0.066325 | 49 | 0.000225 | 0.073419 |
| 21 | 0.084868 | 0.075327 | 51 | 0.013536 | 0.070587 |
| 23 | -0.030260 | 0.068023 | 53 | 0.009802 | 0.082681 |
| 25 | -0.071331 | 0.075356 | 55 | 0.001741 | 0.075473 |
| 27 | -0.023548 | 0.068734 | 57 | -0.004037 | 0.084908 |
| 29 | 0.036678 | 0.074598 | 59 | -0.009270 | 0.083134 |

Table 24. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 4.0$, $r = r' = 16$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.498105 | 0.010361 | 31 | 0.043678 | 0.094802 |
| 3 | -0.440930 | 0.025854 | 33 | 0.054192 | 0.090445 |
| 5 | -0.386977 | 0.045546 | 35 | -0.033193 | 0.102903 |
| 7 | 0.327465 | 0.048362 | 37 | -0.045904 | 0.092539 |
| 9 | 0.275090 | 0.065593 | 39 | 0.014858 | 0.107051 |
| 11 | -0.237449 | 0.069801 | 41 | 0.035231 | 0.091005 |
| 13 | -0.197022 | 0.078309 | 43 | -0.002151 | 0.108170 |
| 15 | 0.175663 | 0.078848 | 45 | -0.031480 | 0.081694 |
| 17 | 0.147499 | 0.086534 | 47 | -0.002831 | 0.103153 |
| 19 | -0.128566 | 0.078821 | 49 | 0.022956 | 0.080287 |
| 21 | -0.114150 | 0.091029 | 51 | 0.003442 | 0.103588 |
| 23 | 0.092986 | 0.083192 | 53 | -0.020424 | 0.081224 |
| 25 | 0.085574 | 0.094227 | 55 | -0.001849 | 0.097451 |
| 27 | -0.065089 | 0.089157 | 57 | 0.020938 | 0.086475 |
| 29 | -0.068886 | 0.091905 | 59 | -0.004269 | 0.087750 |

Table 25. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 2.0$, $r = r' = 20$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.800913 | 0.007078 | 31 | -0.343543 | 0.131223 |
| 3 | 0.400945 | 0.020970 | 33 | -0.194329 | 0.144136 |
| 5 | -0.000350 | 0.034370 | 35 | 0.013523 | 0.139479 |
| 7 | -0.381971 | 0.041943 | 37 | 0.188387 | 0.126604 |
| 9 | -0.620490 | 0.036764 | 39 | 0.279446 | 0.140574 |
| 11 | -0.591510 | 0.061185 | 41 | 0.260172 | 0.153426 |
| 13 | -0.335774 | 0.084348 | 43 | 0.145472 | 0.163489 |
| 15 | 0.009993 | 0.093632 | 45 | -0.015704 | 0.153757 |
| 17 | 0.314036 | 0.073194 | 47 | -0.146161 | 0.144529 |
| 19 | 0.477000 | 0.070349 | 49 | -0.209463 | 0.151062 |
| 21 | 0.446005 | 0.096767 | 51 | -0.195506 | 0.169755 |
| 23 | 0.255641 | 0.116238 | 53 | -0.114079 | 0.168699 |
| 25 | -0.011504 | 0.119011 | 55 | 0.003776 | 0.158401 |
| 27 | -0.251325 | 0.111980 | 57 | 0.104756 | 0.147143 |
| 29 | -0.368194 | 0.112607 | 59 | 0.156100 | 0.158691 |

Table 26. Values of the Correlation Function from
Simulation for $\lambda = 5.0$, $\lambda' = 4.0$,
 $r = 30$, $r' = 16$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.582829 | 0.008997 | 31 | 0.147949 | 0.119429 |
| 3 | -0.227186 | 0.020920 | 33 | -0.062687 | 0.102108 |
| 5 | -0.584209 | 0.024860 | 35 | -0.162847 | 0.120988 |
| 7 | -0.185544 | 0.045243 | 37 | -0.037264 | 0.109908 |
| 9 | 0.421982 | 0.052857 | 39 | 0.112338 | 0.106276 |
| 11 | 0.376335 | 0.064892 | 41 | 0.099340 | 0.122872 |
| 13 | -0.144207 | 0.065912 | 43 | -0.038920 | 0.105417 |
| 15 | -0.383074 | 0.072679 | 45 | -0.116672 | 0.117066 |
| 17 | -0.104752 | 0.080770 | 47 | -0.031048 | 0.126706 |
| 19 | 0.266357 | 0.087604 | 49 | 0.086205 | 0.115937 |
| 21 | 0.238488 | 0.105166 | 51 | 0.075663 | 0.131743 |
| 23 | -0.098663 | 0.085782 | 53 | -0.027275 | 0.117144 |
| 25 | -0.254962 | 0.110398 | 55 | -0.079976 | 0.131977 |
| 27 | -0.056162 | 0.102253 | 57 | -0.029618 | 0.126317 |
| 29 | 0.176781 | 0.112174 | 59 | 0.047288 | 0.122039 |

Table 27. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 4.0$, $r = 64$,
 $r' = 16$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.685654 | 0.012682 | 31 | -0.230835 | 0.022286 |
| 3 | 0.077777 | 0.032270 | 33 | -0.170393 | 0.057682 |
| 5 | -0.215459 | 0.018218 | 35 | -0.033018 | 0.085171 |
| 7 | -0.246131 | 0.015145 | 37 | 0.161748 | 0.087177 |
| 9 | -0.248133 | 0.015372 | 39 | 0.29840 | 0.092775 |
| 11 | -0.248529 | 0.015126 | 41 | 0.276581 | 0.103240 |
| 13 | -0.233263 | 0.020581 | 43 | 0.116649 | 0.105727 |
| 15 | -0.121854 | 0.049227 | 45 | -0.057564 | 0.080740 |
| 17 | 0.162225 | 0.064441 | 47 | -0.171793 | 0.052242 |
| 19 | 0.423126 | 0.080777 | 49 | -0.210108 | 0.037261 |
| 21 | 0.392436 | 0.068165 | 51 | -0.189744 | 0.061055 |
| 23 | 0.136683 | 0.078460 | 53 | -0.117050 | 0.095845 |
| 25 | -0.090888 | 0.062128 | 55 | 0.010766 | 0.109851 |
| 27 | -0.204828 | 0.031683 | 57 | 0.146534 | 0.098255 |
| 29 | -0.237740 | 0.016807 | 59 | 0.216715 | 0.100968 |

Table 28. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 10.0$, $r = 50$,
 $r' = 30$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.466413 | 0.009918 | 31 | 0.211283 | 0.114067 |
| 3 | -0.480892 | 0.018080 | 33 | 0.173172 | 0.124005 |
| 5 | -0.445639 | 0.024912 | 35 | -0.186913 | 0.121309 |
| 7 | 0.420443 | 0.041828 | 37 | -0.151670 | 0.122505 |
| 9 | 0.383115 | 0.055692 | 39 | 0.160441 | 0.133336 |
| 11 | -0.375723 | 0.047766 | 41 | 0.134025 | 0.131237 |
| 13 | -0.341901 | 0.064561 | 43 | -0.141010 | 0.135912 |
| 15 | 0.337105 | 0.068660 | 45 | -0.114693 | 0.134955 |
| 17 | 0.295318 | 0.087183 | 47 | 0.115359 | 0.144711 |
| 19 | -0.301711 | 0.072537 | 49 | 0.098216 | 0.141384 |
| 21 | -0.257538 | 0.094141 | 51 | -0.103725 | 0.153155 |
| 23 | 0.266840 | 0.091081 | 53 | -0.082500 | 0.143143 |
| 25 | 0.224671 | 0.107403 | 55 | 0.083229 | 0.151007 |
| 27 | -0.241301 | 0.095979 | 57 | 0.069009 | 0.146872 |
| 29 | -0.198852 | 0.110255 | 59 | -0.071706 | 0.155844 |

Table 28. Values of the Correlation Function from
Simulation for $\lambda = \lambda' = 5.0$,
 $r = r' = 50$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.799033 | 0.005422 | 31 | -0.545380 | 0.093508 |
| 3 | 0.398790 | 0.014833 | 33 | -0.299323 | 0.114783 |
| 5 | -0.000388 | 0.023119 | 35 | 0.037208 | 0.113904 |
| 7 | -0.396415 | 0.030444 | 37 | 0.349083 | 0.096362 |
| 9 | -0.715076 | 0.032848 | 39 | 0.524309 | 0.093468 |
| 11 | -0.690560 | 0.041146 | 41 | 0.493777 | 0.119141 |
| 13 | -0.369174 | 0.051994 | 43 | 0.268317 | 0.142709 |
| 15 | 0.017951 | 0.057852 | 45 | -0.045881 | 0.137478 |
| 17 | 0.387639 | 0.056090 | 47 | -0.327981 | 0.114287 |
| 19 | 0.634053 | 0.049784 | 49 | -0.483659 | 0.114190 |
| 21 | 0.605248 | 0.070384 | 51 | -0.447528 | 0.145068 |
| 23 | 0.334392 | 0.083959 | 53 | -0.235001 | 0.162415 |
| 25 | -0.028717 | 0.087633 | 55 | 0.052458 | 0.154846 |
| 27 | -0.369685 | 0.076894 | 57 | 0.307516 | 0.128713 |
| 29 | -0.576205 | 0.069951 | 59 | 0.438259 | 0.130335 |

Table 30. Values of the Correlation Function from
Simulation for $\mu = \mu' = 5.0$, $\sigma^2 = \sigma'^2 = 6.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.623522 | 0.018319 | 31 | 0.010179 | 0.070706 |
| 3 | -0.002097 | 0.050690 | 33 | 0.007010 | 0.074460 |
| 5 | -0.304172 | 0.047961 | 35 | -0.011380 | 0.076358 |
| 7 | -0.226272 | 0.071736 | 37 | -0.025597 | 0.077356 |
| 9 | -0.000821 | 0.081884 | 39 | -0.026064 | 0.079132 |
| 11 | 0.122845 | 0.065977 | 41 | -0.000538 | 0.072645 |
| 13 | 0.087293 | 0.071628 | 45 | 0.015304 | 0.074056 |
| 15 | -0.003082 | 0.071139 | 45 | 0.009238 | 0.070067 |
| 17 | -0.040723 | 0.070065 | 47 | -0.002811 | 0.069950 |
| 19 | -0.021067 | 0.079806 | 49 | -0.010254 | 0.065992 |
| 21 | 0.009259 | 0.071174 | 51 | -0.015300 | 0.069669 |
| 23 | 0.008291 | 0.075548 | 53 | -0.009892 | 0.071907 |
| 25 | -0.003669 | 0.076870 | 55 | -0.000529 | 0.067384 |
| 27 | -0.018472 | 0.075723 | 57 | 0.009735 | 0.078840 |
| 29 | -0.006276 | 0.071537 | 59 | -0.006554 | 0.079604 |

Table 31. Values of the Correlation Function from
Simulation for $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 6.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.803596 | 0.009320 | 31 | -0.352753 | 0.119012 |
| 3 | 0.409491 | 0.027803 | 33 | -0.249968 | 0.141182 |
| 5 | 0.018160 | 0.043627 | 35 | -0.066490 | 0.144691 |
| 7 | -0.333957 | 0.049695 | 37 | 0.117361 | 0.139338 |
| 9 | -0.564641 | 0.044596 | 39 | 0.245013 | 0.130270 |
| 11 | -0.588024 | 0.055694 | 41 | 0.270201 | 0.136066 |
| 13 | -0.372786 | 0.094737 | 45 | 0.191976 | 0.146822 |
| 15 | -0.055000 | 0.111716 | 45 | 0.051367 | 0.152767 |
| 17 | 0.241952 | 0.099460 | 47 | -0.093272 | 0.147608 |
| 19 | 0.429348 | 0.084332 | 49 | -0.192403 | 0.141651 |
| 21 | 0.447417 | 0.091409 | 51 | -0.211141 | 0.150950 |
| 23 | 0.304888 | 0.122288 | 53 | -0.159432 | 0.168111 |
| 25 | 0.061878 | 0.144596 | 55 | -0.052790 | 0.166366 |
| 27 | -0.175675 | 0.137688 | 57 | 0.058798 | 0.158792 |
| 29 | -0.329146 | 0.112784 | 59 | 0.136816 | 0.157828 |

Table 32. Values of the Correlation Function from
Simulation for $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 4.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.803537 | 0.007190 | 31 | -0.448293 | 0.106695 |
| 3 | 0.409618 | 0.021153 | 33 | -0.298980 | 0.145328 |
| 5 | 0.013423 | 0.034752 | 35 | -0.052956 | 0.166544 |
| 7 | -0.363154 | 0.041905 | 37 | 0.189469 | 0.156441 |
| 9 | -0.629743 | 0.040212 | 39 | 0.352990 | 0.124673 |
| 11 | -0.642759 | 0.047004 | 41 | 0.377225 | 0.125276 |
| 13 | -0.395778 | 0.083870 | 43 | 0.252252 | 0.159463 |
| 15 | -0.035343 | 0.097793 | 45 | 0.050379 | 0.183540 |
| 17 | 0.298531 | 0.091331 | 47 | 0.161063 | 0.178518 |
| 19 | 0.514067 | 0.074617 | 49 | -0.298814 | 0.154094 |
| 21 | 0.535011 | 0.077588 | 51 | -0.320998 | 0.147776 |
| 23 | 0.345904 | 0.117320 | 53 | -0.230933 | 0.177518 |
| 25 | 0.049567 | 0.139059 | 55 | -0.054341 | 0.201585 |
| 27 | -0.242847 | 0.131227 | 57 | 0.123501 | 0.193313 |
| 29 | -0.430594 | 0.102560 | 59 | 0.240878 | 0.171849 |

Table 33. Values of the Correlation Function from
Simulation for $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 2.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.801897 | 0.002082 | 31 | -0.527470 | 0.043756 |
| 3 | 0.403519 | 0.006056 | 33 | -0.331005 | 0.060914 |
| 5 | 0.002932 | 0.009012 | 35 | -0.027545 | 0.057055 |
| 7 | -0.393859 | 0.010089 | 37 | 0.230987 | 0.038965 |
| 9 | -0.709090 | 0.009966 | 39 | 0.398755 | 0.029757 |
| 11 | -0.724183 | 0.020453 | 41 | 0.440105 | 0.052670 |
| 13 | -0.397223 | 0.022710 | 45 | 0.296918 | 0.075225 |
| 15 | 0.001536 | 0.024433 | 45 | 0.063849 | 0.073993 |
| 17 | 0.380232 | 0.027720 | 47 | -0.160427 | 0.050275 |
| 19 | 0.609775 | 0.015578 | 49 | -0.314864 | 0.037098 |
| 21 | 0.613435 | 0.028853 | 51 | -0.364762 | 0.059718 |
| 23 | 0.372464 | 0.041662 | 53 | -0.264572 | 0.085273 |
| 25 | -0.000925 | 0.044785 | 55 | -0.088006 | 0.083257 |
| 27 | -0.312224 | 0.036855 | 57 | 0.088736 | 0.058782 |
| 29 | -0.499362 | 0.025684 | 59 | 0.039049 | 0.039820 |

Table 34. Values of the Correlation Function from
Simulation for $\mu = 4.0$, $\mu' = 20.0$,
 $\sigma^2 = 1.0$, $\sigma'^2 = 4.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.700410 | 0.019123 | 31 | -0.152586 | 0.037872 |
| 3 | 0.124412 | 0.044020 | 33 | -0.189395 | 0.018743 |
| 5 | -0.168071 | 0.018020 | 35 | -0.193176 | 0.016242 |
| 7 | -0.191685 | 0.016395 | 37 | -0.192012 | 0.018595 |
| 9 | -0.191609 | 0.013838 | 39 | -0.193694 | 0.022826 |
| 11 | -0.192520 | 0.014585 | 41 | -0.147460 | 0.046052 |
| 13 | -0.193444 | 0.014818 | 43 | -0.037343 | 0.082201 |
| 15 | -0.193527 | 0.016571 | 45 | 0.148818 | 0.119778 |
| 17 | -0.181645 | 0.023304 | 47 | 0.305118 | 0.122522 |
| 19 | -0.099709 | 0.048574 | 49 | 0.315341 | 0.105509 |
| 21 | 0.141687 | 0.078999 | 51 | 0.200265 | 0.128167 |
| 23 | 0.423242 | 0.076919 | 53 | 0.047716 | 0.133527 |
| 25 | 0.452245 | 0.076361 | 55 | -0.086890 | 0.083692 |
| 27 | 0.225756 | 0.092043 | 57 | -0.162081 | 0.045187 |
| 29 | -0.024955 | 0.075477 | 59 | -0.187465 | 0.024652 |

Table 35. Values of the Correlation Function from
Simulation for $\mu = 30.0$, $\mu' = 5.0$,
 $\sigma^2 = \sigma'^2 = 4.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.770224 | 0.021801 | 31 | 0.116626 | 0.091423 |
| 3 | 0.337499 | 0.048864 | 33 | 0.376190 | 0.097041 |
| 5 | 0.018078 | 0.045482 | 35 | 0.507910 | 0.092307 |
| 7 | -0.123406 | 0.026791 | 37 | 0.433925 | 0.110801 |
| 9 | -0.159727 | 0.018457 | 39 | 0.234206 | 0.123240 |
| 11 | -0.165282 | 0.018778 | 41 | 0.040857 | 0.092242 |
| 13 | -0.166457 | 0.019094 | 43 | -0.083258 | 0.053029 |
| 15 | -0.167467 | 0.019144 | 45 | -0.144822 | 0.028425 |
| 17 | -0.168410 | 0.019207 | 47 | -0.164140 | 0.022683 |
| 19 | -0.169132 | 0.019194 | 49 | -0.167282 | 0.020609 |
| 21 | -0.169640 | 0.019338 | 51 | -0.168863 | 0.020548 |
| 23 | -0.169668 | 0.019347 | 53 | -0.169866 | 0.020648 |
| 25 | -0.170183 | 0.019680 | 55 | -0.170575 | 0.020480 |
| 27 | -0.160366 | 0.025139 | 57 | -0.170989 | 0.020326 |
| 29 | -0.089816 | 0.056328 | 59 | -0.166934 | 0.024395 |

Table 36. Values of the Correlation Function from
Simulation for $\mu = \mu' = 10.0$,
 $\sigma^2 = 2.0$, $\sigma'^2 = 0.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.802929 | 0.003255 | 31 | -0.688596 | 0.060403 |
| 3 | 0.407293 | 0.009496 | 33 | -0.408092 | 0.090811 |
| 5 | 0.009872 | 0.015129 | 35 | -0.025639 | 0.103237 |
| 7 | -0.388680 | 0.019508 | 37 | 0.351509 | 0.100674 |
| 9 | -0.751575 | 0.020958 | 39 | 0.629652 | 0.070965 |
| 11 | -0.762707 | 0.025227 | 41 | 0.651329 | 0.078634 |
| 13 | -0.410723 | 0.036652 | 43 | 0.400817 | 0.116995 |
| 15 | -0.015305 | 0.044003 | 45 | 0.031427 | 0.130553 |
| 17 | 0.382239 | 0.048452 | 47 | -0.335833 | 0.123096 |
| 19 | 0.707204 | 0.042861 | 48 | -0.600913 | 0.084903 |
| 21 | 0.724663 | 0.046645 | 51 | -0.626324 | 0.089905 |
| 23 | 0.414218 | 0.065017 | 53 | -0.391693 | 0.135627 |
| 25 | 0.020371 | 0.073812 | 55 | -0.036863 | 0.155366 |
| 27 | -0.367333 | 0.074220 | 57 | 0.318631 | 0.145819 |
| 29 | -0.669097 | 0.053310 | 59 | 0.570497 | 0.102714 |

Table 37. Values of the Correlation Function from
Simulation for $\mu = 5.0$, $\mu' = 30.0$,
 $\sigma^2 = 1.0$, $\sigma'^2 = 4.0$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_u |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.770422 | 0.011290 | 31 | 0.145574 | 0.089411 |
| 3 | 0.311345 | 0.032440 | 33 | 0.452071 | 0.096976 |
| 5 | -0.064296 | 0.026593 | 35 | 0.595551 | 0.079876 |
| 7 | -0.162740 | 0.009619 | 37 | 0.464721 | 0.111688 |
| 9 | -0.165329 | 0.009892 | 39 | 0.182725 | 0.104930 |
| 11 | -0.166473 | 0.010169 | 41 | -0.038276 | 0.062538 |
| 13 | -0.167501 | 0.010151 | 43 | -0.137183 | 0.028612 |
| 15 | -0.168163 | 0.010150 | 45 | -0.161049 | 0.013242 |
| 17 | -0.168906 | 0.010315 | 47 | -0.165749 | 0.009756 |
| 19 | -0.169958 | 0.010413 | 49 | -0.166899 | 0.009667 |
| 21 | -0.171020 | 0.010371 | 51 | -0.167750 | 0.009905 |
| 23 | -0.171659 | 0.010365 | 53 | -0.168772 | 0.010153 |
| 25 | -0.171083 | 0.011564 | 55 | -0.169804 | 0.010232 |
| 27 | -0.158926 | 0.019553 | 57 | -0.168924 | 0.014066 |
| 29 | -0.081326 | 0.050973 | 59 | -0.164469 | 0.023020 |

Table 38. Values of the Correlation Function from
Simulation for $\mu = 4.0$, $\mu' = 20.0$,
 $\sigma^2 = \sigma'^2 = 0.25$

| TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U | TIME INTERVAL | SIM.AVE. SER.CORR. | STD.DEV. OF ρ_U |
|------------------|-----------------------|-------------------------|------------------|-----------------------|-------------------------|
| 1 | 0.701098 | 0.009493 | 31 | -0.198340 | 0.011635 |
| 3 | 0.103945 | 0.028592 | 33 | -0.197389 | 0.011360 |
| 5 | -0.196487 | 0.010431 | 35 | -0.198471 | 0.007736 |
| 7 | -0.196008 | 0.007844 | 37 | -0.197719 | 0.008427 |
| 9 | -0.194979 | 0.009912 | 39 | -0.197136 | 0.010364 |
| 11 | -0.195870 | 0.008875 | 41 | -0.197224 | 0.008056 |
| 13 | -0.196122 | 0.009158 | 43 | -0.184614 | 0.015407 |
| 15 | -0.196308 | 0.011784 | 45 | 0.093581 | 0.079035 |
| 17 | -0.196689 | 0.008414 | 47 | 0.627309 | 0.076558 |
| 19 | -0.196552 | 0.007094 | 49 | 0.657727 | 0.085479 |
| 21 | 0.089122 | 0.036238 | 51 | 0.151610 | 0.099731 |
| 23 | 0.669413 | 0.040278 | 53 | -0.159989 | 0.034506 |
| 25 | 0.690339 | 0.053659 | 55 | -0.197195 | 0.013284 |
| 27 | 0.124097 | 0.061287 | 57 | -0.197413 | 0.009478 |
| 29 | -0.182542 | 0.016318 | 59 | -0.197472 | 0.007753 |

Table 39. Analysis of Natural Populations

```

2 COMMENT ANALYSIS OF NATURAL POPULATIONS
2 INTEGER J, I, K, M, C, SC, N, F, L, F, TOP, INTV
2 ARRAY X(2500), A(100), AP(100), WV(480), F(50)
2 INPUT DATA1(C, M, SC)
2 INPUT DATA2(FOR J=1,1,M) SA(1)
2 INPUT DATA3(FOR I=1,1,M) SAP(1)
2 OUTPUT RF(1)(C, R)
2 OUTPUT RF3(1)(K, WV(1)) $OUTPUT RF4(F(1), WV(1))
2 OUTPUT RF2(1)(I, 1), AD(1)
2 FORMAT HEAD(R20, *CTHRY NO, *, 13, *CORRELATION A, A PRIME, R2, V10, F, W2)
2 FORMAT HEAD(R36, *A, R4, *A DRIME, W2)
2 FORMAT HEAD(R30, *SERIAL CORRELATION FUNCTION, W2)
2 FORMAT HEAD(R10, *K, R15, *RHO(K), R10, W2)
2 FORMAT HEAD(R30, *Q(X) AS A FUNCTION OF X, W2)
2 FORMAT C4(10, 13, R10, X10, 6, W0) $FORMAT HEAD(R10, *K, R10, *Q(F), W2)
2 FORMAT CPY2(130, 13, X10, 2, X10, 2, W0)
2 OUTPUT CYL(CYC(F))
2 FORMAT COPCYL(10, *THE MEAN CYCLE LENGTH I, X10, 5, W2)
2 OUTPUT TEST(FOR I=1,1,1920) X(1)
2 FORMAT CORRECT(10, 20X4, 1, W0)
2 OUTPUT DOT(D)
2 FORMAT COPDOT(R10, *THE VALUE OF P, X10, 7, W0)
2 OUTPUT PHN(FOR I=1,1,1920) X(1)
2 FORMAT CORPHN(10R(*5, 15X4, 1, W0))
2 F(1)=25F(2)=35F(3)=45F(4)=105 F(5)=55 F(6)=125 F(7)=155F(8)=16
2 F(9)=205 F(10)=245 F(11)=305 F(12)=325 F(13)=405 F(14)=485F(15)=64
2 F(16)=805 F(17)=905 F(18)=120 F(19)=455F(20)=8
2 P1,, READ(SSDATA1)
2 READ(SSDATA2)
2 READ(SSDATA3)
2 CA=0.0 S=0.0
2 CAP=0.0
2 CAA=0.0
2 CAD=0.0
2 CAAP=0.0
2 FOR I=1,1,M)
2 BEGIN CAA=CA+AP(I)
2 CAP=CAP+AP(I)
2 CAA=CAA+AP(I)*2
2 CAD=CAD+AP(I)*2
2 CAAP=CAAP+(A(1), AP(I)) FND
2 R=((M, CAAP)-(CA, CAP)/SQRT((M, CAA-(CA*2))(M, CCAP-(CAP*2)))
2 S=CA+CAP
2 IF S NFO 480.0
2 BEGIN IF S LSS 480.0
2 BEGIN D=480.0-S
2 A=(M+A)/M+D FND
2 D= S-480.0
2 A(M)=A(M)-D FND
2 CYCLF=(480.0)/FLOAT(M)
2 D2,, CC=0.0
2 J=1
2 FOR I=1,1,M)
2 BEGIN CAA=CAA+AP(I)
2 UNTIL (J), (0, 25) GTO CC
2 BEGIN X(J)=1.0
2 J=J+1 FND
2 X(J)=0.0
2 CAA=CAA+AP(I)
2 UNTIL (J), (0, 25) GTO CC
2 BEGIN X(J)=0.0
2 J=J+1 FND
2 X(J)=1.0 FND
2 WRITE(SSPUN, COPPHN) $CTOP
2 WRITE(SSDFC), HEAD)
2 WRITE(SSCYL, COPCYL)
2 WRITE(SSHEAD2)
2 FOR I=1,1,M)
2 WRITE(SSDFC2, CPY2)
2 SUM=0.0
2 FOR I=1,1,1920)
2 SUM=SUM+X(I)
2 P=(SUM)/(1920.0)
2 V=P(1)-P)
2 WRITE(SSDOT, COPDOT)
2 WRITE(SSHEAD5) $WRITE(SSHEAD6)
2 FOR L=1,1,20)
2 BEGIN SR=0.0 S=480.0 S=SSQ=0.0 $INTV=F(1)
2 FOR K=1,1,F(L)) $RFGIN TOP=1920-F(L)+K
2 FOR J=1K, INTV, TOP)

```

Table 39. (Continued)

```

2          S7=C2+X(J)
2          D=(S3/1920.0)/FLOAT(FIL1)
2          CQ=CQ+(11/P)/(1.0-P)/(1920.0)/FLOAT(FIL1)
2          S7=0.0      FND
2          RV=(1920.0/FLOAT(FIL1))-(1+CQ/(V.FIL1))
2          WVIL=SQRT(RV)
2          WRITE($$RE4,C4)  $  S50=0.0      FND
2          WRITE($$HEAD3)
2          WRITE($$HEAD4)
2          FOR J=1,1,1920) $X(1)=X(J)-D
2          FOR K=1,1,260)
2          BEGIN S=0.0
2          FOR I=1,1,(1920-K))
2          S=S+(X(I),X(I+K))
2          WV(K)=S/(V*(1920-K))
2          WRITE($$RE34,C4) FND
2          *TOP $ FINISH

```

Table 40. Extended Analysis of Simulation Data

```

2  COMMENT DATA ANALYSIS HISTORGRAM AND SUMMARY HINES
2  INTEGER J,K,L,M,N,P
2  ARRAY X(20,50),AVF(50),STD(50),UR(10),N(10)
2  INPUT DATA(M,7)
2  OUTPUT RESULT(AVF(J),STD(J))
2  FORMAT COPY(*AVF(J)=*,X10.6,P6,*STD(J)=*,X10.6,W0)
2  OUTPUT MAX|MIN|I(J,MA,M)IS FORMAT COPY(*J=*,I3,P5,*MA=*,X10.6,
2  P5,*M|=*,X10.6,W0) $ OUTPUT HISTO(FOR I=1,1,7)S(UR(I),N(I))
2  FORMAT COP(7(RI),X10.6,P5,(3,W0))
2  FOR K=(1,1,50)
2  BEGIN FOR J=1,1,18)
2  BEGIN READ($$DATA)
2  X(J,K)=Z      FND FND
2  FOR J=1,1,18)
2  BEGIN C=C+X(J,K)
2  C=C+X(J,K)*2      FND
2  AVF(J)=C/50.0
2  STD(J)=SQRT(((C5)-(150.0)(AVF(J)*2)))/(49.0)
2  MA=MAX(X(J,1),X(J,2))
2  MI=MIN(X(J,1),X(J,2))
2  FOR K=(1,1,50)
2  BEGIN MA=MAX(MA,X(J,K)) MI=MIN(MI,X(J,K)) FND
2  WRITE($$MAX|MIN|I,COPYM) $WRITE($$P5|J|I,COPYV)
2  SP=MA-MI*0.000007
2  CIW=SP/7.0
2  UR(I)=MI-0.000005
2  FOR I=(2,1,7) $UR(I)=UR(I)+(I-1).CIW
2  FOR I=(1,1,7) $N(I)=0 $ FOR I=(1,1,7)
2  BEGIN FOR K=(1,1,50)
2  BEGIN IF X(J,K) GT= UR(I)
2  N(I)=N(I)+1 $ XXX=0.0      FND FND
2  WRITE($$HISTO,COP) FND
2  STOP
2  FINISH

```

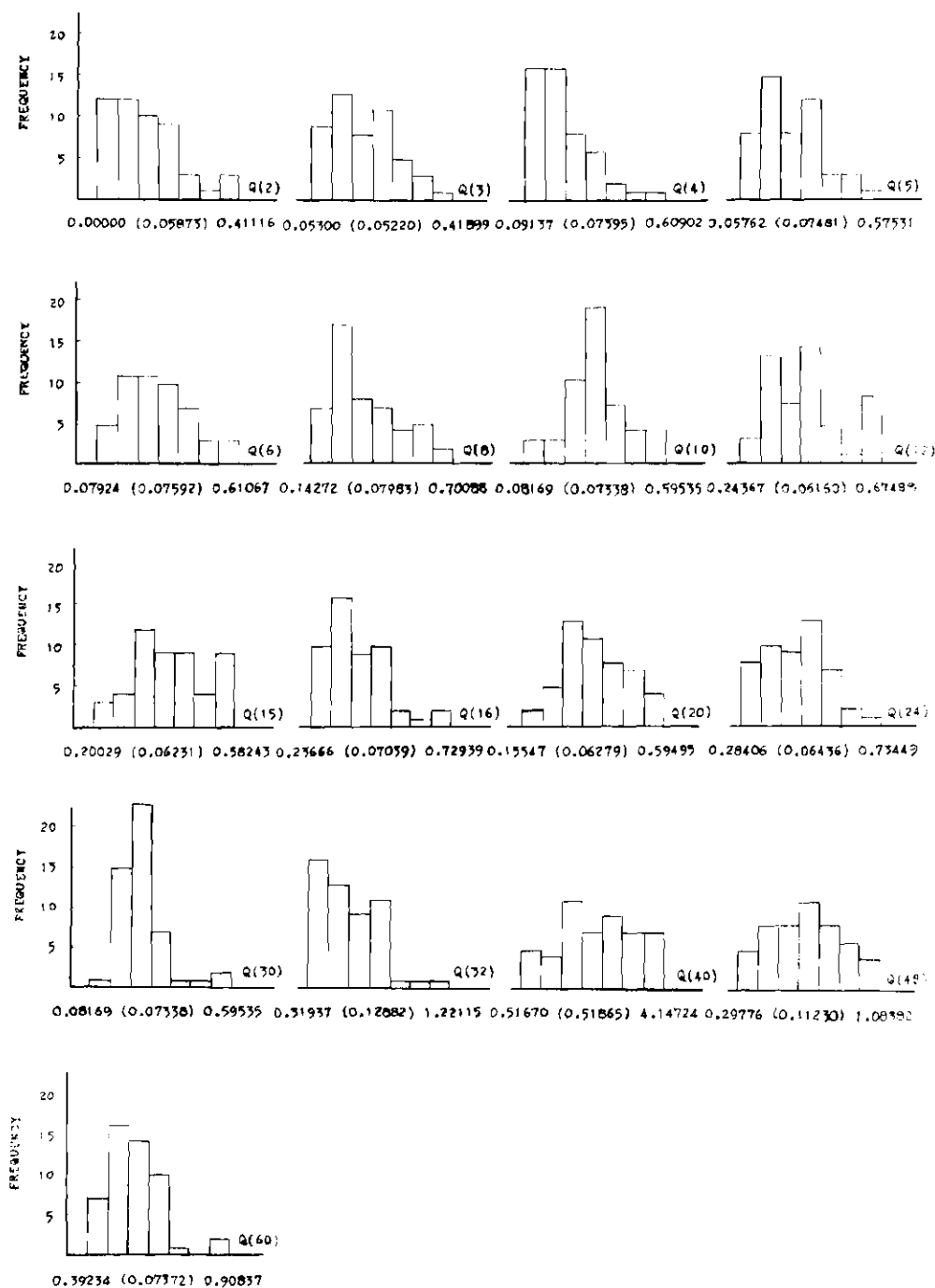


Figure 37. Histograms for $Q(k)$ from Simulations with Gamma Distributed Span Lengths, $\lambda = \lambda' = 5.0$, and $r = r' = 50$ (Given Are Initial and Final Points with Interval Width in Parentheses)

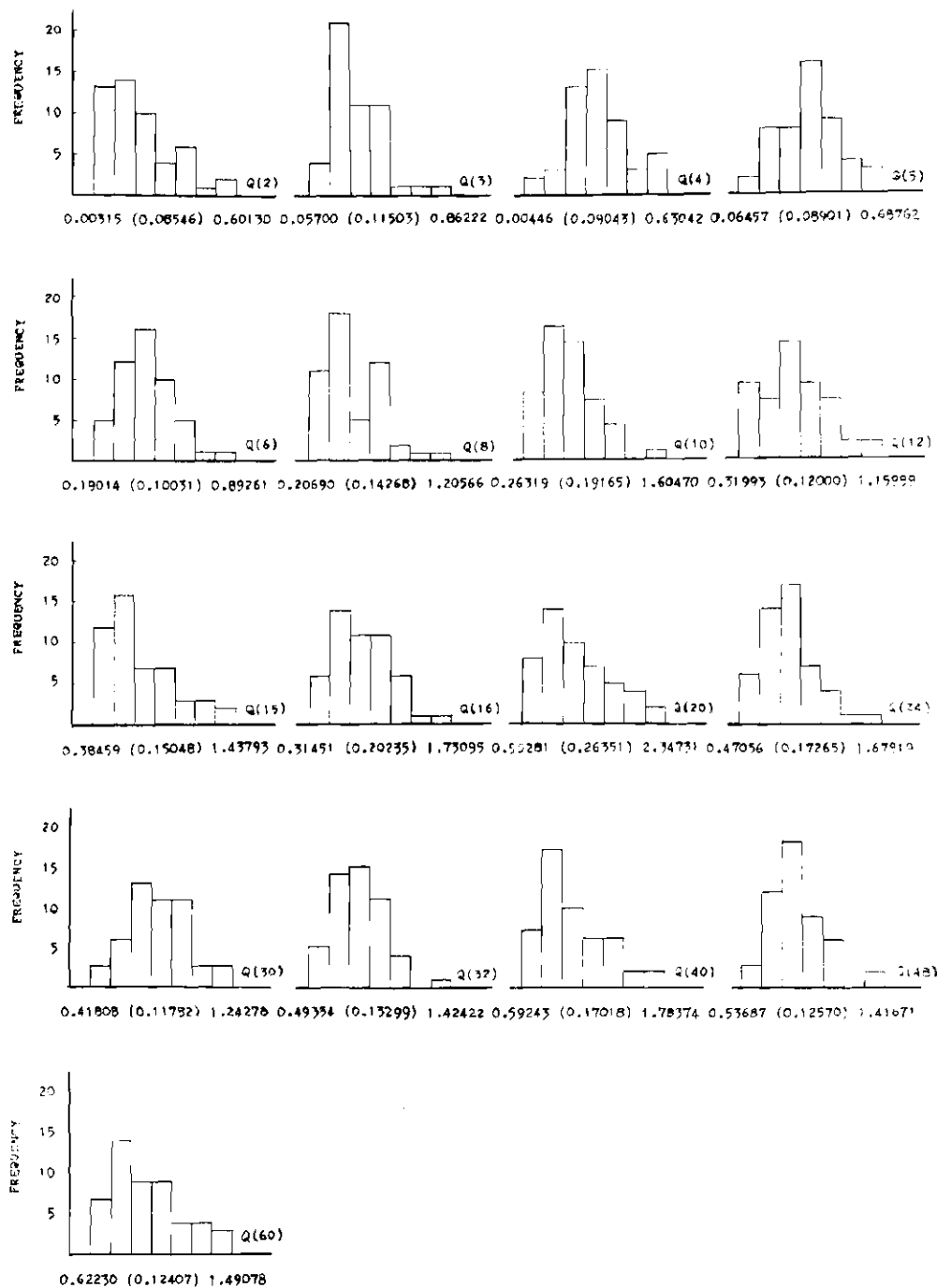


Figure 38. Histograms for $Q(k)$ from Simulations with Gamma Distributed Span Lengths, $\lambda = 3.0$, $\lambda' = 2.0$, $r = 9$, and $r' = 14$ (Given are Initial and Final Points with Interval Width in Parentheses)

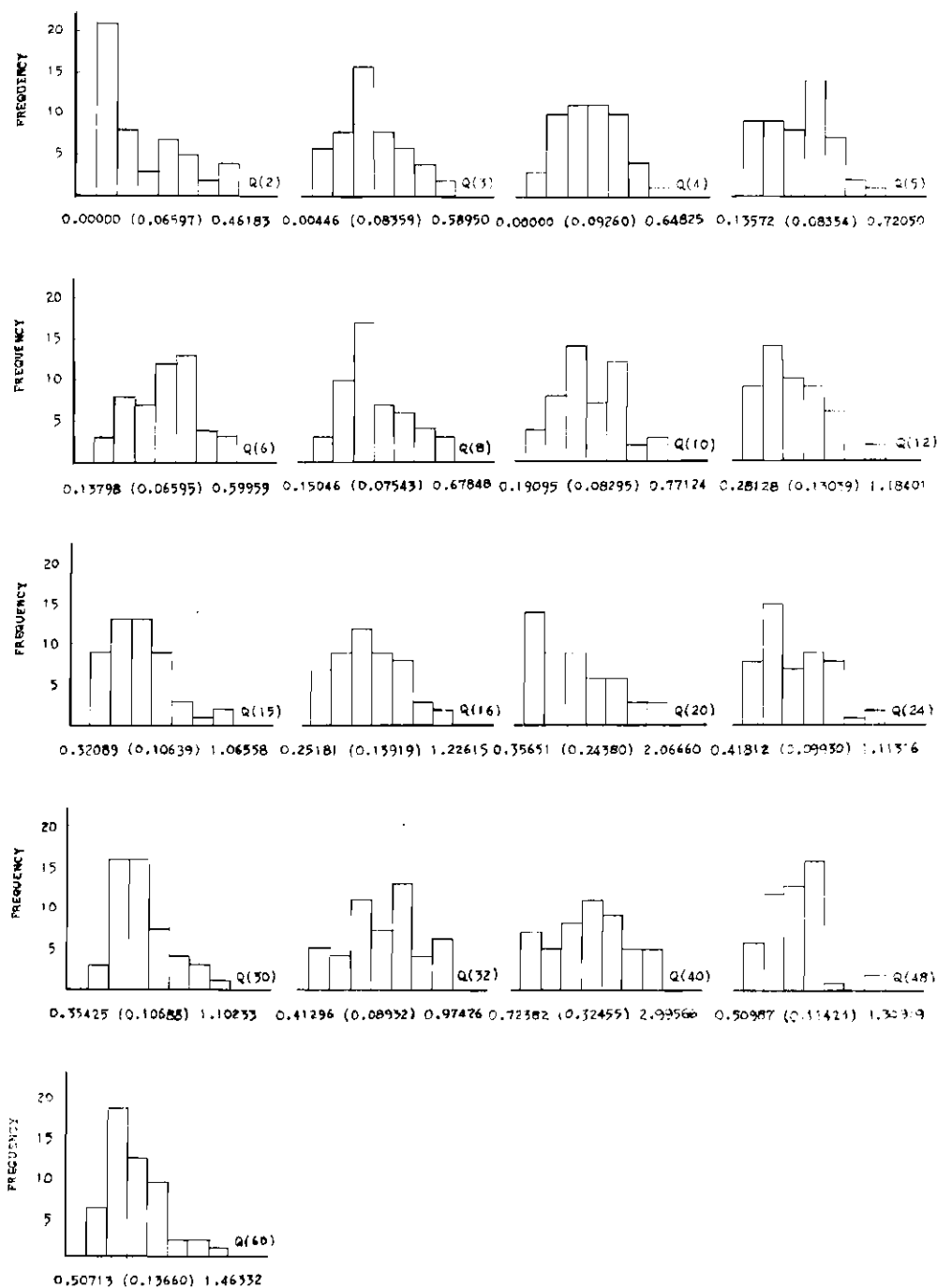


Figure 39. Histograms for $Q(k)$ from Simulations with Gamma Distributed Span Lengths, $\lambda = \lambda' = 4.0$, $r = 64$, $r' = 16$ (Given are Initial and Final Points with Interval Width in Parentheses)

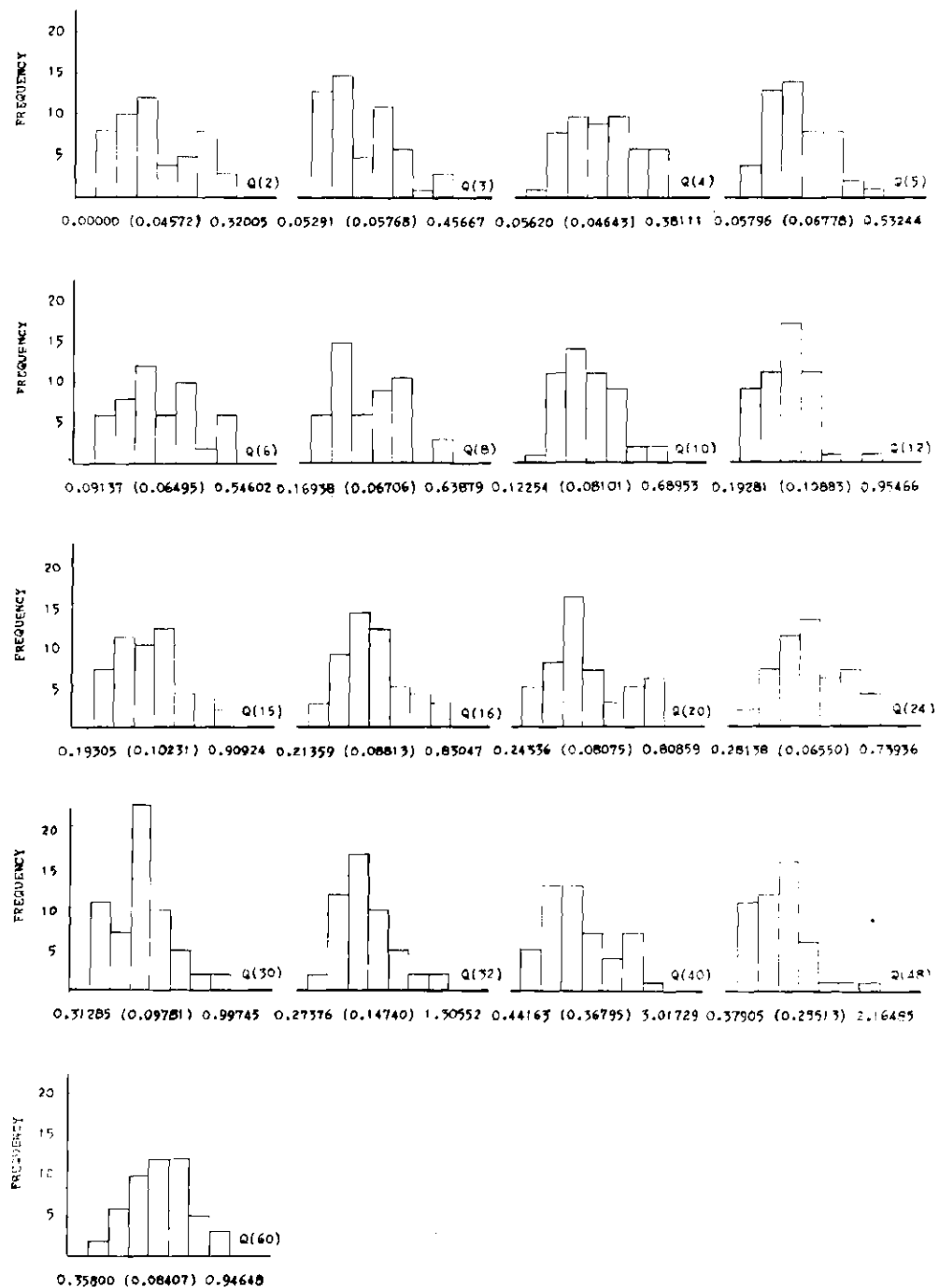


Figure 40. Histograms for $Q(k)$ from Simulation with Truncated Normal Distributions of Span Length $\mu = \mu' = 10.0$, $\sigma^2 = \sigma'^2 = 6.0$ (Given Are Initial and Final Points with Interval Width in Parentheses)

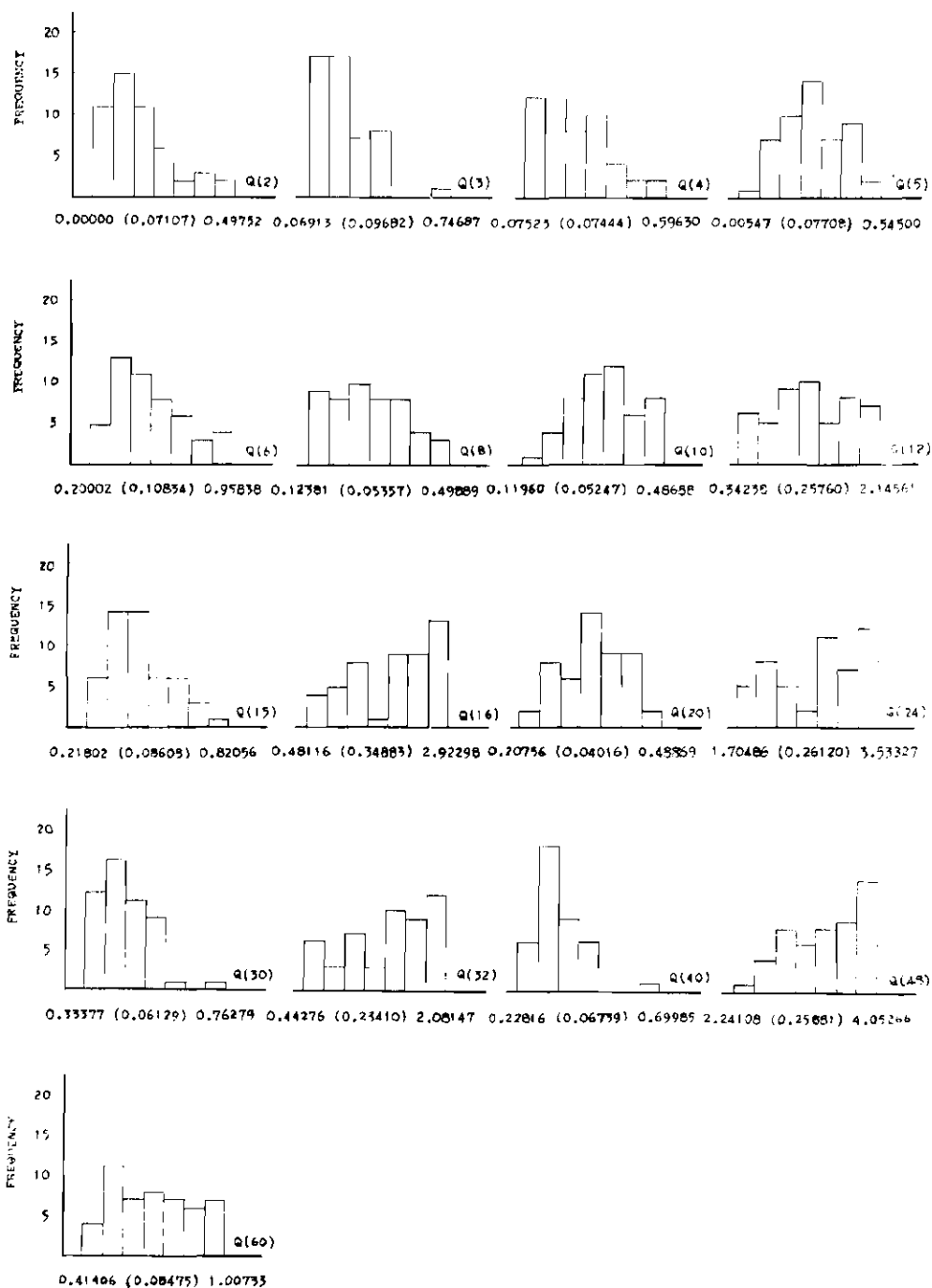


Figure 41. Histograms for $Q(k)$ from Simulation with Truncated Normal Distributions of Span Length $\mu = 4.0$, $\mu' = 20.0$, $\sigma^2 = \sigma'^2 = 0.25$ (Given Are Initial and Final Points with Interval Width in Parentheses)

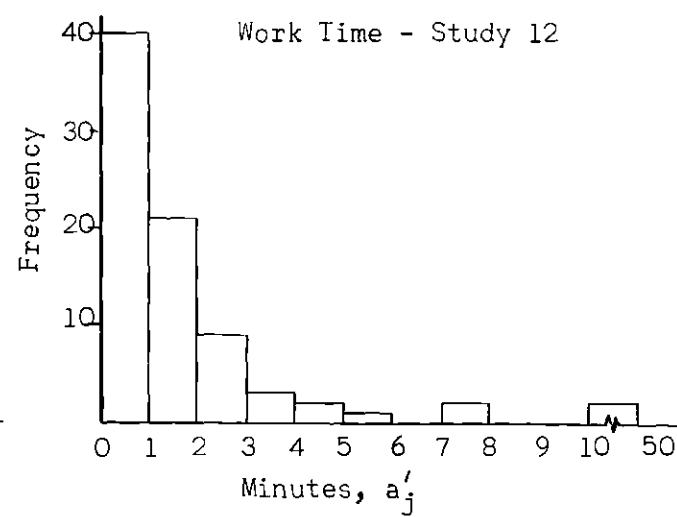
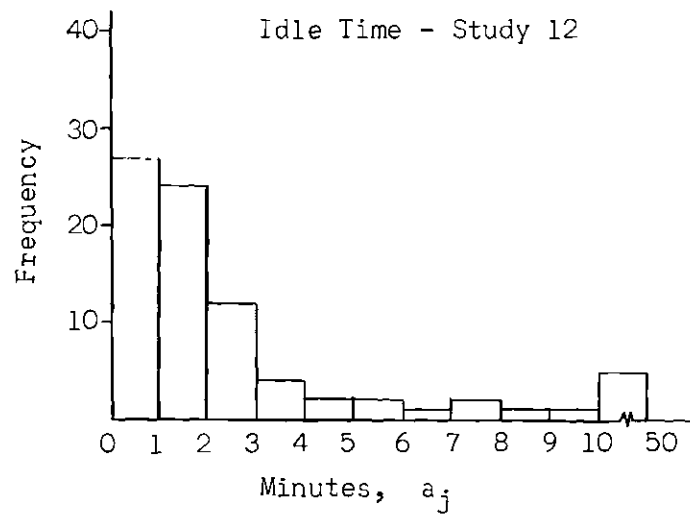
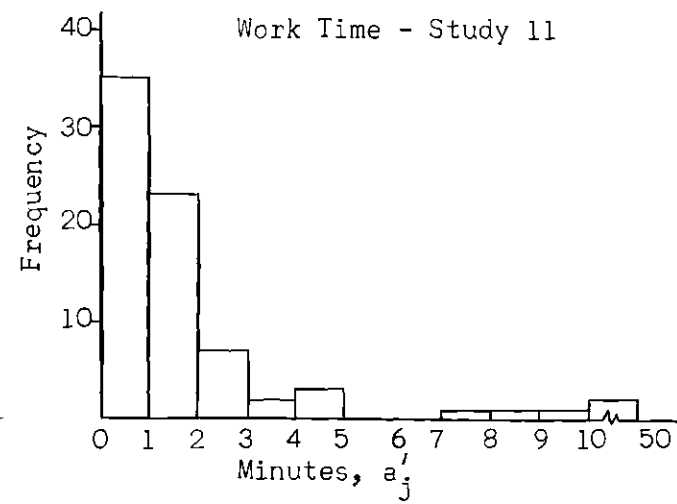
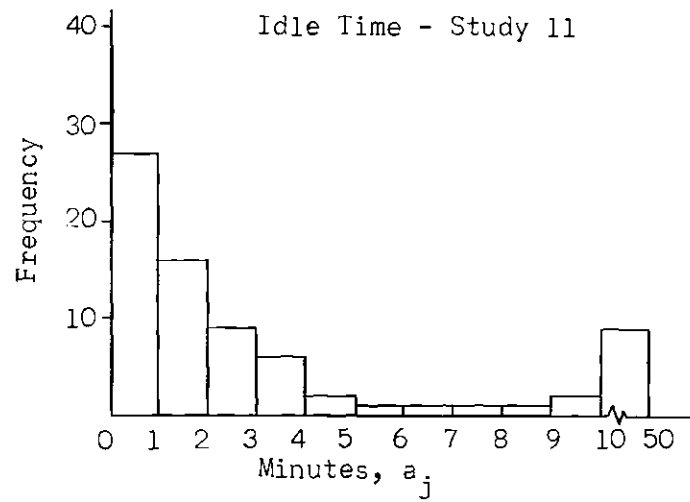


Figure 42. Histograms of Work and Idle Time

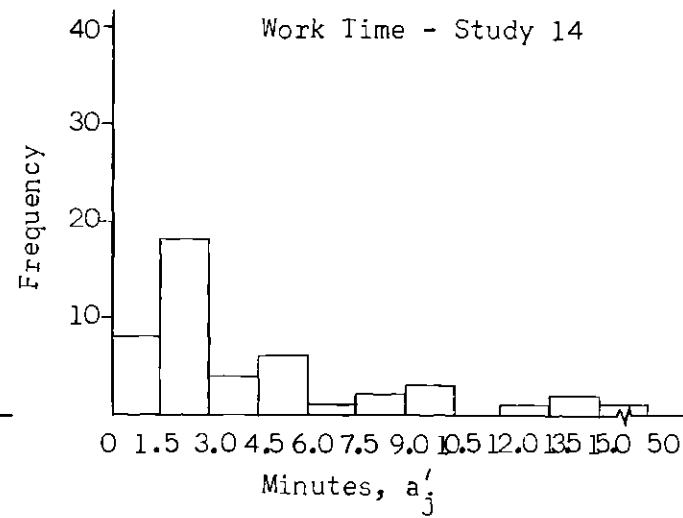
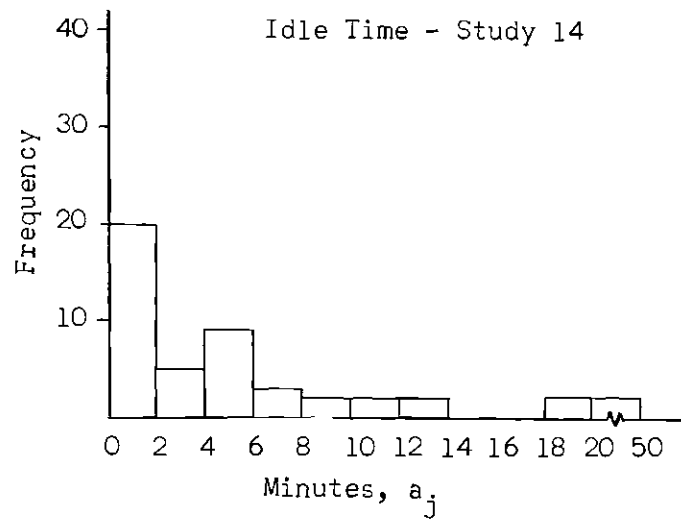
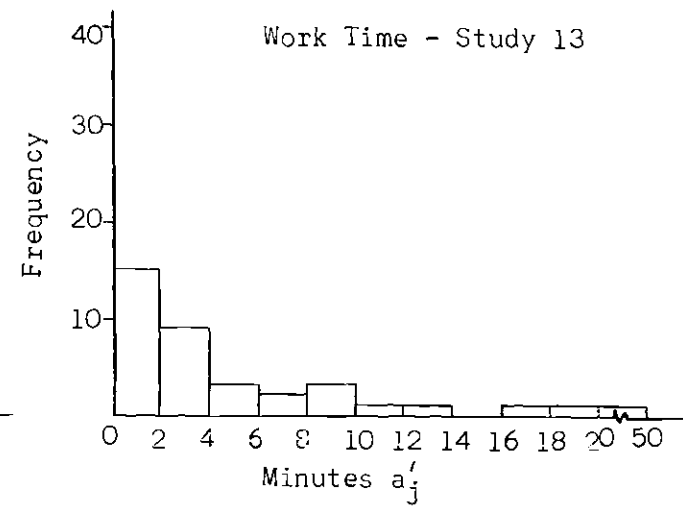
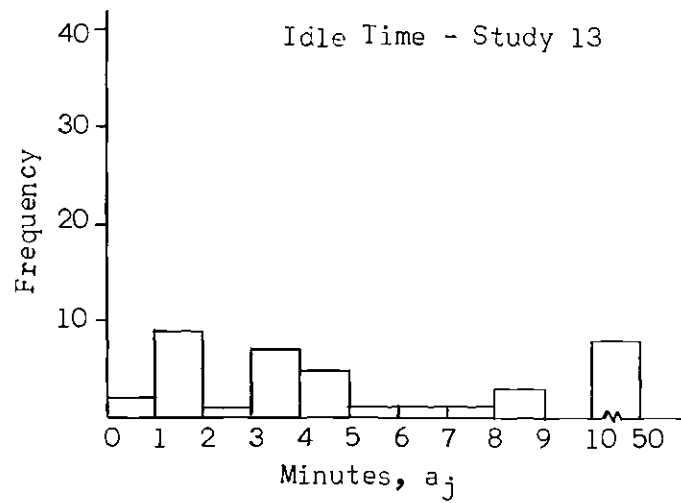


Figure 42. (Continued)

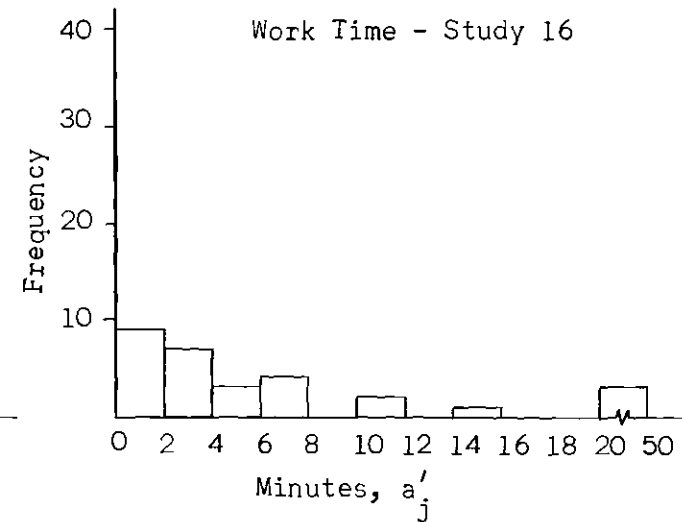
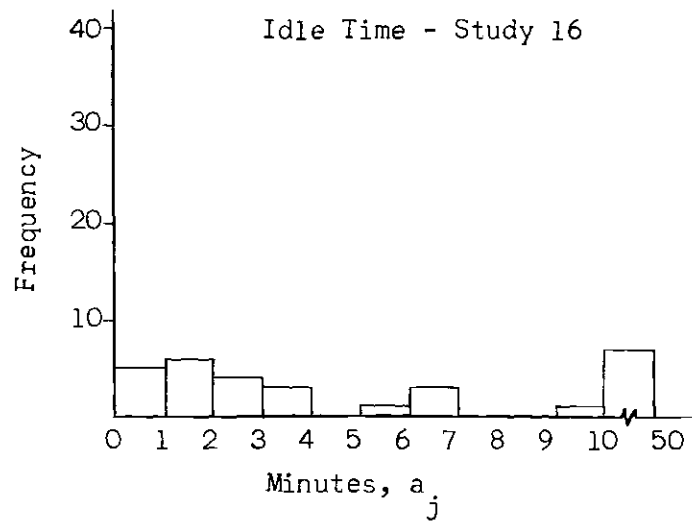
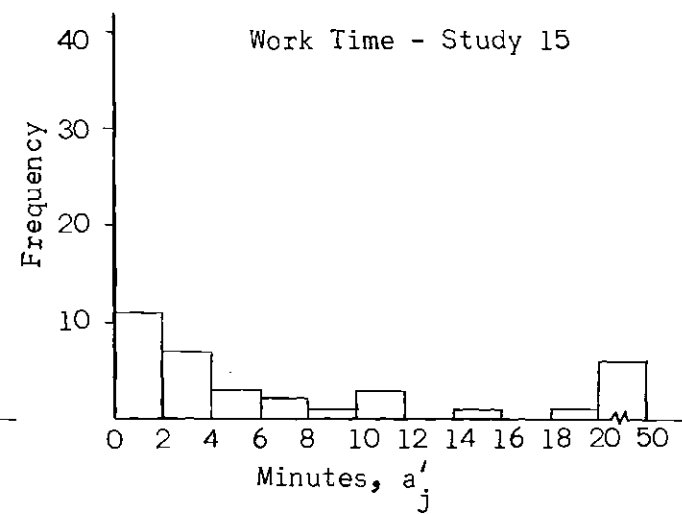
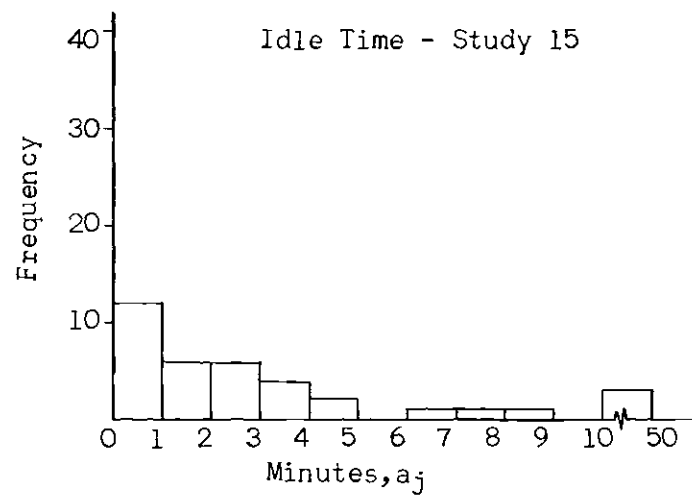


Figure 42. (Continued)

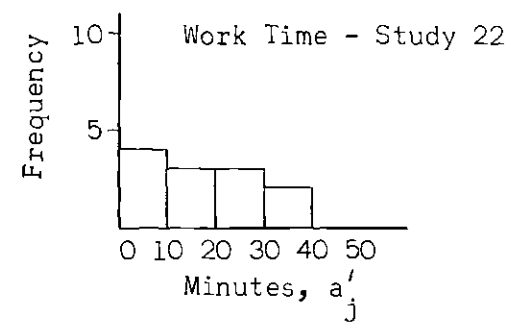
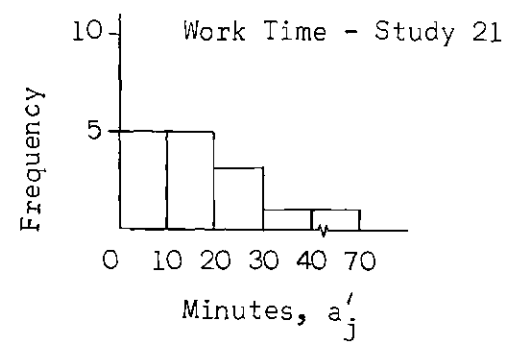
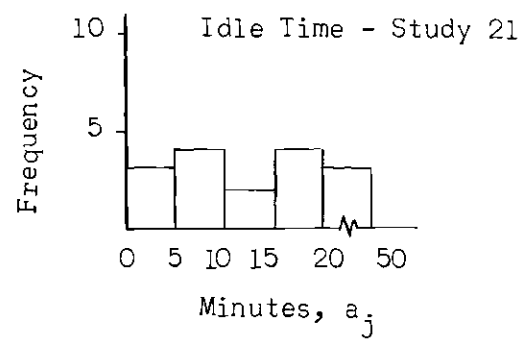


Figure 42. (Continued)

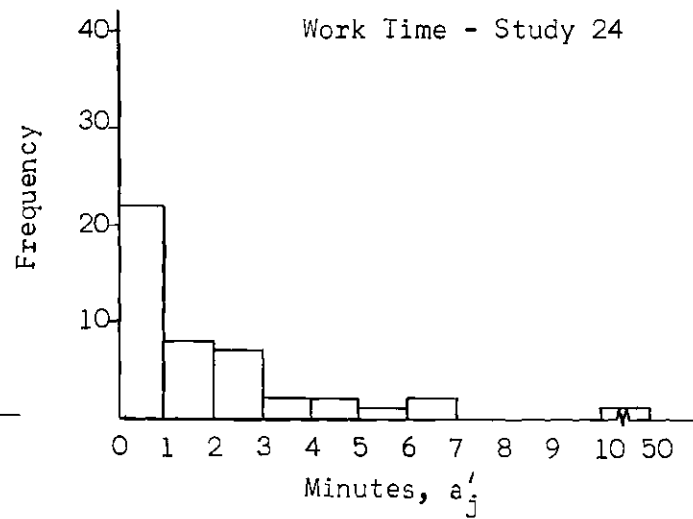
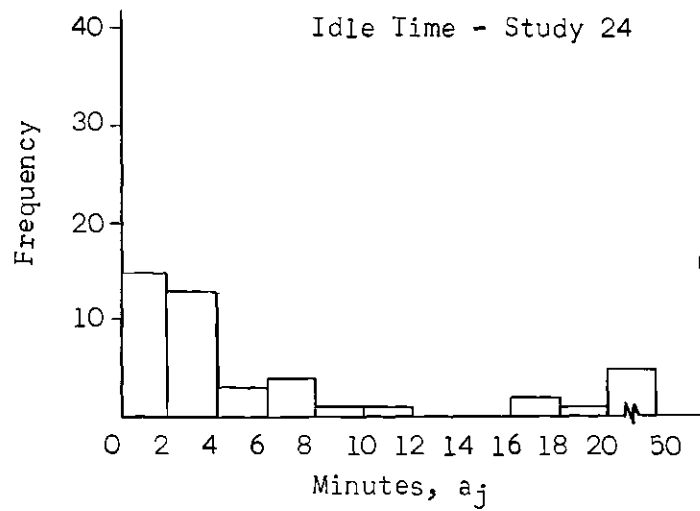
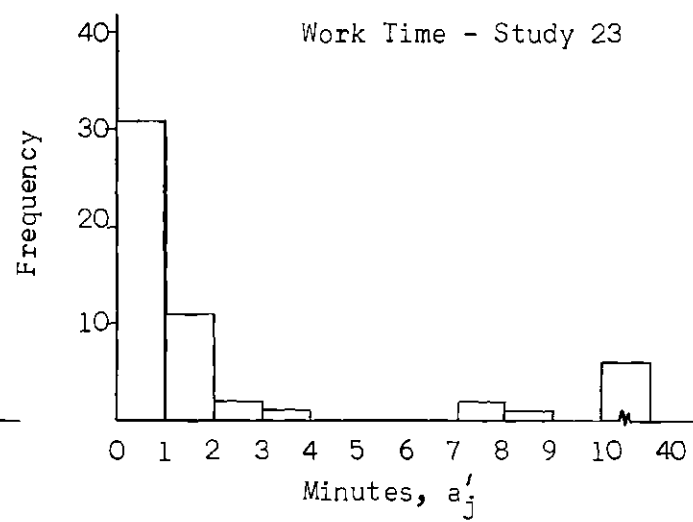
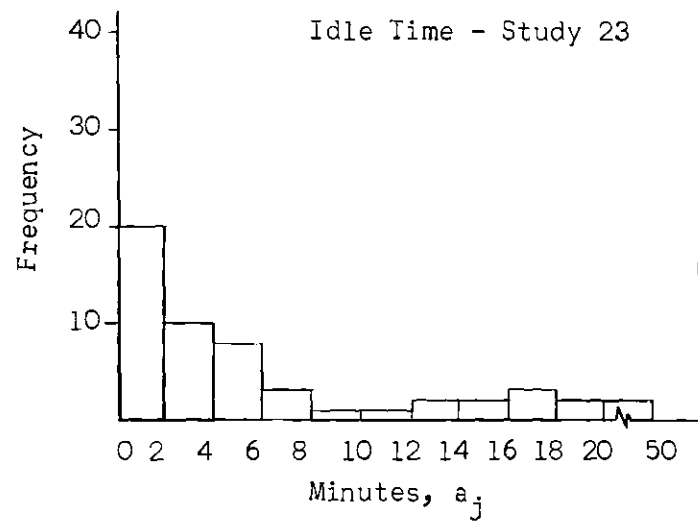


Figure 42. (Continued)

Table 42. Data from Study No. 11 (Span Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|-----------------|-----------------------|-----------------------|
| 1 | 13.81 | 0.49 | 39 | 0.59 | 1.60 |
| 2 | 3.86 | 1.26 | 40 | 3.44 | 0.77 |
| 3 | 1.65 | 0.30 | 41 | 0.65 | 2.00 |
| 4 | 1.12 | 0.72 | 42 | 0.65 | 0.35 |
| 5 | 3.50 | 1.13 | 43 | 0.36 | 3.62 |
| 6 | 3.51 | 0.21 | 44 | 0.38 | 3.00 |
| 7 | 0.54 | 0.95 | 45 | 1.19 | 0.18 |
| 8 | 2.39 | 2.67 | 46 | 1.63 | 9.90 |
| 9 | 1.19 | 1.36 | 47 | 0.56 | 1.12 |
| 10 | 0.72 | 1.47 | 48 | 0.36 | 1.39 |
| 11 | 0.43 | 4.15 | 49 | 48.46 | 0.51 |
| 12 | 0.74 | 0.35 | 50 | 0.21 | 0.67 |
| 13 | 1.24 | 0.25 | 51 | 2.44 | 0.85 |
| 14 | 4.56 | 0.67 | 52 | 0.28 | 1.02 |
| 15 | 3.69 | 1.63 | 53 | 1.45 | 0.15 |
| 16 | 0.60 | 2.51 | 54 | 20.72 | 1.66 |
| 17 | 0.96 | 0.29 | 55 | 0.48 | 0.31 |
| 18 | 1.60 | 1.28 | 56 | 0.64 | 1.39 |
| 19 | 1.60 | 0.21 | 57 | 1.43 | 2.09 |
| 20 | 0.93 | 0.33 | 58 | 4.85 | 2.76 |
| 21 | 1.65 | 0.66 | 59 | 1.94 | 0.87 |
| 22 | 2.28 | 1.25 | 60 | 7.36 | 0.46 |
| 23 | 1.55 | 0.64 | 61 | 10.00 | 1.49 |
| 24 | 6.87 | 7.13 | 62 | 0.25 | 0.56 |
| 25 | 0.35 | 1.38 | 63 | 8.81 | 8.29 |
| 26 | 15.30 | 2.95 | 64 | 16.13 | 1.09 |
| 27 | 1.10 | 1.22 | 65 | 17.74 | 15.41 |
| 28 | 0.62 | 1.53 | 66 | 2.30 | 3.97 |
| 29 | 1.06 | 2.66 | 67 | 2.25 | 0.36 |
| 30 | 2.98 | 0.40 | 68 | 2.05 | 0.48 |
| 31 | 1.96 | 2.51 | 69 | 0.82 | 0.86 |
| 32 | 10.84 | 26.86 | 70 | 3.17 | 0.48 |
| 33 | 3.00 | 0.70 | 71 | 0.76 | 0.56 |
| 34 | 0.97 | 0.31 | 72 | 0.15 | 0.62 |
| 35 | 5.61 | 0.46 | 73 | 11.15 | 1.42 |
| 36 | 9.30 | 4.94 | 74 | 0.22 | 1.49 |
| 37 | 2.26 | 4.05 | 75 | 30.60 | 0.00 |
| 38 | 0.35 | 1.24 | | | |

Table 43. Data from Study No. 12 (Span
Duration in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|-----------------|-----------------------|-----------------------|
| 1 | 8.00 | 0.80 | 42 | 28.10 | 2.25 |
| 2 | 3.75 | 2.37 | 43 | 1.85 | 0.90 |
| 3 | 1.75 | 5.59 | 44 | 2.96 | 0.70 |
| 4 | 0.93 | 0.21 | 45 | 2.02 | 1.69 |
| 5 | 0.38 | 0.82 | 46 | 0.41 | 1.12 |
| 6 | 0.93 | 0.46 | 47 | 1.90 | 1.19 |
| 7 | 0.92 | 1.21 | 48 | 1.18 | 0.30 |
| 8 | 0.61 | 0.70 | 49 | 2.85 | 1.80 |
| 9 | 0.43 | 1.13 | 50 | 2.04 | 0.82 |
| 10 | 0.34 | 4.53 | 51 | 1.15 | 0.41 |
| 11 | 0.32 | 2.07 | 52 | 4.70 | 0.40 |
| 12 | 6.15 | 1.24 | 53 | 1.49 | 0.35 |
| 13 | 0.63 | 0.80 | 54 | 44.10 | 0.83 |
| 14 | 0.84 | 1.25 | 55 | 4.01 | 0.82 |
| 15 | 3.07 | 0.58 | 56 | 2.75 | 0.39 |
| 16 | 3.18 | 1.03 | 57 | 16.48 | 0.75 |
| 17 | 1.88 | 1.25 | 58 | 1.50 | 0.20 |
| 18 | 0.55 | 0.94 | 59 | 1.66 | 2.83 |
| 19 | 1.40 | 1.62 | 60 | 1.05 | 3.21 |
| 20 | 0.65 | 0.79 | 61 | 0.77 | 2.35 |
| 21 | 0.50 | 1.50 | 62 | 3.93 | 40.74 |
| 22 | 1.58 | 0.54 | 63 | 10.50 | 1.23 |
| 23 | 0.90 | 0.76 | 64 | 2.55 | 0.96 |
| 24 | 0.35 | 0.08 | 65 | 1.60 | 2.67 |
| 25 | 2.24 | 0.70 | 66 | 0.78 | 1.32 |
| 26 | 1.35 | 1.12 | 67 | 2.20 | 1.15 |
| 27 | 0.37 | 38.56 | 68 | 1.42 | 3.21 |
| 28 | 1.34 | 0.49 | 69 | 0.85 | 2.70 |
| 29 | 2.23 | 0.20 | 70 | 9.80 | 1.05 |
| 30 | 1.40 | 0.28 | 71 | 0.85 | 0.90 |
| 31 | 1.45 | 0.45 | 72 | 5.37 | 2.80 |
| 32 | 1.58 | 1.15 | 73 | 5.07 | 1.58 |
| 33 | 0.75 | 1.96 | 74 | 0.67 | 3.02 |
| 34 | 0.95 | 0.59 | 75 | 1.45 | 1.21 |
| 35 | 7.42 | 7.70 | 76 | 1.65 | 0.95 |
| 36 | 1.07 | 7.70 | 77 | 1.00 | 0.68 |
| 37 | 0.84 | 4.75 | 78 | 2.60 | 0.34 |
| 38 | 0.64 | 0.94 | 79 | 1.75 | 0.21 |
| 39 | 2.20 | 0.66 | 80 | 1.40 | 2.65 |
| 40 | 8.66 | 0.30 | 81 | 32.55 | 0.00 |
| 41 | 2.18 | 0.25 | | | |

Table 44. Data from Study No. 13 (Span
Duration in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|-----------------|-----------------------|-----------------------|
| 1 | 4.00 | 3.00 | 20 | 4.88 | 0.87 |
| 2 | 3.90 | 2.95 | 21 | 4.06 | 42.21 |
| 3 | 1.78 | 16.14 | 22 | 47.63 | 12.11 |
| 4 | 1.95 | 9.54 | 23 | 8.20 | 2.80 |
| 5 | 6.81 | 0.72 | 24 | 7.67 | 6.23 |
| 6 | 3.18 | 18.90 | 25 | 1.61 | 5.43 |
| 7 | 3.09 | 5.54 | 26 | 11.12 | 2.61 |
| 8 | 4.80 | 0.72 | 27 | 8.23 | 3.18 |
| 9 | 2.42 | 1.70 | 28 | 1.60 | 0.31 |
| 10 | 14.10 | 3.61 | 29 | 25.76 | 0.61 |
| 11 | 1.07 | 7.38 | 30 | 10.86 | 8.64 |
| 12 | 1.24 | 0.66 | 31 | 3.02 | 1.07 |
| 13 | 0.77 | 1.08 | 32 | 14.51 | 0.58 |
| 14 | 1.74 | 3.39 | 33 | 4.83 | 1.07 |
| 15 | 1.83 | 2.32 | 34 | 0.64 | 5.72 |
| 16 | 8.18 | 3.61 | 35 | 4.48 | 1.77 |
| 17 | 3.55 | 8.58 | 36 | 1.43 | 0.53 |
| 18 | 17.89 | 0.78 | 37 | 3.53 | 11.98 |
| 19 | 5.17 | 1.09 | 38 | 29.04 | 0.00 |

Table 45. Data from Study No. 14 (Span Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|-----------------|-----------------------|-----------------------|
| 1 | 4.01 | 1.98 | 25 | 1.63 | 3.41 |
| 2 | 2.13 | 2.35 | 26 | 1.10 | 1.51 |
| 3 | 11.74 | 2.95 | 27 | 1.05 | 1.29 |
| 4 | 5.62 | 14.11 | 28 | 4.23 | 1.35 |
| 5 | 12.09 | 5.33 | 29 | 5.02 | 9.71 |
| 6 | 3.25 | 1.74 | 30 | 5.08 | 0.27 |
| 7 | 2.55 | 9.00 | 31 | 0.46 | 6.61 |
| 8 | 6.81 | 1.99 | 32 | 0.70 | 3.87 |
| 9 | 0.35 | 1.75 | 33 | 8.21 | 24.12 |
| 10 | 19.15 | 2.53 | 34 | 41.46 | 9.49 |
| 11 | 1.35 | 2.98 | 35 | 1.95 | 0.48 |
| 12 | 2.07 | 13.39 | 36 | 0.51 | 1.52 |
| 13 | 18.29 | 2.52 | 37 | 5.87 | 1.61 |
| 14 | 13.19 | 8.35 | 38 | 0.68 | 2.85 |
| 15 | 7.08 | 5.54 | 39 | 0.55 | 2.96 |
| 16 | 0.31 | 5.72 | 40 | 8.67 | 4.38 |
| 17 | 0.51 | 7.33 | 41 | 0.95 | 2.31 |
| 18 | 0.60 | 14.83 | 42 | 7.36 | 1.11 |
| 19 | 2.03 | 4.96 | 43 | 1.07 | 4.91 |
| 20 | 0.90 | 2.93 | 44 | 1.74 | 1.33 |
| 21 | 0.54 | 1.32 | 45 | 4.51 | 1.57 |
| 22 | 1.51 | 9.50 | 46 | 5.70 | 2.30 |
| 23 | 0.36 | 3.71 | 47 | 10.69 | 0.00 |
| 24 | 24.14 | 4.46 | | | |

Table 46. Data from Study No. 15 (Span Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|--------------|--------------------|--------------------|--------------|--------------------|--------------------|
| 1 | 3.90 | 3.37 | 19 | 6.90 | 5.94 |
| 2 | 0.43 | 46.75 | 20 | 1.20 | 0.83 |
| 3 | 2.30 | 11.12 | 21 | 2.65 | 14.15 |
| 4 | 0.60 | 0.33 | 22 | 0.80 | 0.42 |
| 5 | 2.00 | 0.70 | 23 | 38.93 | 39.07 |
| 6 | 2.40 | 9.00 | 24 | 2.08 | 3.10 |
| 7 | 0.40 | 4.00 | 25 | 8.65 | 28.80 |
| 8 | 3.55 | 4.05 | 26 | 0.43 | 1.12 |
| 9 | 1.30 | 1.52 | 27 | 0.90 | 7.28 |
| 10 | 0.55 | 29.73 | 28 | 0.77 | 6.23 |
| 11 | 0.80 | 21.15 | 29 | 2.78 | 0.75 |
| 12 | 0.85 | 1.30 | 30 | 3.82 | 5.95 |
| 13 | 4.60 | 2.70 | 31 | 13.35 | 1.05 |
| 14 | 7.50 | 10.05 | 32 | 2.07 | 0.60 |
| 15 | 1.70 | 2.12 | 33 | 0.98 | 11.95 |
| 16 | 3.67 | 0.48 | 34 | 1.40 | 19.93 |
| 17 | 0.58 | 3.68 | 35 | 1.37 | 20.38 |
| 18 | 4.71 | 3.54 | 36 | 25.94 | 0.00 |

Table 47. Data from Study No. 16 (Span Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|--------------|--------------------|--------------------|--------------|--------------------|--------------------|
| 1 | 2.50 | 2.25 | 16 | 2.60 | 1.55 |
| 2 | 1.47 | 0.48 | 17 | 11.70 | 14.80 |
| 3 | 3.25 | 5.90 | 18 | 2.65 | 2.20 |
| 4 | 0.27 | 11.88 | 19 | 9.73 | 39.07 |
| 5 | 6.42 | 5.73 | 20 | 46.55 | 25.60 |
| 6 | 3.57 | 2.35 | 21 | 1.02 | 6.83 |
| 7 | 1.83 | 4.69 | 22 | 5.75 | 10.90 |
| 8 | 1.26 | 0.93 | 23 | 0.60 | 1.30 |
| 9 | 0.90 | 0.57 | 24 | 1.45 | 1.45 |
| 10 | 3.15 | 0.60 | 25 | 1.75 | 7.35 |
| 11 | 16.40 | 7.68 | 26 | 52.10 | 7.87 |
| 12 | 6.42 | 41.74 | 27 | 2.78 | 2.30 |
| 13 | 0.85 | 2.03 | 28 | 25.10 | 3.00 |
| 14 | 6.92 | 1.20 | 29 | 17.12 | 0.55 |
| 15 | 0.75 | 3.80 | 30 | 26.54 | 0.00 |

Table 48. Data from Study No. 21 (Span Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|
| 1 | 1.50 | 8.21 |
| 2 | 42.13 | 38.92 |
| 3 | 19.88 | 1.94 |
| 4 | 14.18 | 10.55 |
| 5 | 8.85 | 3.66 |
| 6 | 5.67 | 31.09 |
| 7 | 8.09 | 14.94 |
| 8 | 15.26 | 28.42 |
| 9 | 38.17 | 27.66 |
| 10 | 41.10 | 6.15 |
| 11 | 40.50 | 29.71 |
| 12 | 8.42 | 12.80 |
| 13 | 22.20 | 0.00 |

Table 49. Data from Study No. 22 (Span Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|
| 1 | 2.50 | 0.50 |
| 2 | 1.95 | 7.95 |
| 3 | 18.99 | 1.40 |
| 4 | 12.16 | 67.39 |
| 5 | 6.07 | 11.41 |
| 6 | 14.85 | 6.00 |
| 7 | 9.22 | 11.15 |
| 8 | 5.83 | 22.42 |
| 9 | 16.86 | 9.75 |
| 10 | 3.68 | 27.70 |
| 11 | 18.51 | 0.91 |
| 12 | 34.50 | 18.82 |
| 13 | 23.45 | 33.76 |
| 14 | 17.67 | 20.20 |
| 15 | 9.48 | 10.54 |
| 16 | 34.38 | 0.00 |

Table 50. Data from Study No. 23 (Span
Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|-----------------|-----------------------|-----------------------|
| 1 | 0.00 | 0.80 | 29 | 0.97 | 3.49 |
| 2 | 6.15 | 0.15 | 30 | 17.21 | 1.24 |
| 3 | 1.60 | 0.10 | 31 | 13.81 | 1.15 |
| 4 | 10.65 | 10.30 | 32 | 17.55 | 0.85 |
| 5 | 16.70 | 1.30 | 33 | 14.40 | 0.88 |
| 6 | 20.60 | 0.30 | 34 | 2.44 | 0.39 |
| 7 | 6.70 | 1.50 | 35 | 1.82 | 1.78 |
| 8 | 46.80 | 14.85 | 36 | 0.54 | 17.88 |
| 9 | 0.91 | 0.44 | 37 | 0.52 | 7.30 |
| 10 | 5.95 | 1.55 | 38 | 1.40 | 1.04 |
| 11 | 0.65 | 0.62 | 39 | 3.97 | 0.27 |
| 12 | 0.68 | 1.10 | 40 | 1.63 | 0.13 |
| 13 | 4.20 | 0.15 | 41 | 2.66 | 0.14 |
| 14 | 2.70 | 0.77 | 42 | 0.96 | 10.69 |
| 15 | 2.32 | 0.78 | 43 | 5.48 | 10.34 |
| 16 | 1.95 | 0.42 | 44 | 4.31 | 0.81 |
| 17 | 4.77 | 0.36 | 45 | 1.31 | 10.42 |
| 18 | 0.18 | 0.37 | 46 | 1.86 | 0.79 |
| 19 | 0.31 | 0.14 | 47 | 1.67 | 0.43 |
| 20 | 19.53 | 0.73 | 48 | 1.98 | 2.12 |
| 21 | 3.02 | 1.53 | 49 | 1.59 | 0.52 |
| 22 | 5.57 | 1.30 | 50 | 14.92 | 0.46 |
| 23 | 8.35 | 0.27 | 51 | 4.41 | 0.49 |
| 24 | 2.38 | 0.50 | 52 | 12.47 | 0.25 |
| 25 | 2.75 | 1.33 | 53 | 7.58 | 1.00 |
| 26 | 1.10 | 8.62 | 54 | 5.12 | 2.73 |
| 27 | 2.65 | 7.49 | 55 | 3.15 | 0.00 |
| 28 | 19.62 | 0.12 | | | |

Table 51. Data from Study No. 24 (Span
Durations in Minutes)

| CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION | CYCLE NUMBER | IDLE SPAN DURATION | WORK SPAN DURATION |
|-----------------|-----------------------|-----------------------|-----------------|-----------------------|-----------------------|
| 1 | 2.08 | 0.80 | 24 | 0.38 | 0.86 |
| 2 | 0.84 | 1.90 | 25 | 33.03 | 0.87 |
| 3 | 1.44 | 0.22 | 26 | 5.86 | 2.34 |
| 4 | 0.64 | 0.48 | 27 | 0.47 | 0.23 |
| 5 | 0.55 | 6.80 | 28 | 1.43 | 2.99 |
| 6 | 1.50 | 28.22 | 29 | 1.09 | 0.44 |
| 7 | 16.21 | 1.97 | 30 | 3.34 | 2.26 |
| 8 | 0.85 | 0.67 | 31 | 19.42 | 2.06 |
| 9 | 6.88 | 0.66 | 32 | 26.24 | 0.75 |
| 10 | 4.13 | 0.67 | 33 | 81.41 | 6.66 |
| 11 | 1.14 | 4.22 | 34 | 5.86 | 1.92 |
| 12 | 20.32 | 2.70 | 35 | 2.02 | 1.74 |
| 13 | 2.31 | 0.21 | 36 | 2.39 | 5.25 |
| 14 | 6.85 | 0.69 | 37 | 1.07 | 2.56 |
| 15 | 0.77 | 0.50 | 38 | 21.94 | 0.58 |
| 16 | 17.34 | 0.19 | 39 | 0.31 | 0.09 |
| 17 | 23.21 | 0.43 | 40 | 7.36 | 3.63 |
| 18 | 0.95 | 0.70 | 41 | 9.77 | 2.59 |
| 19 | 3.37 | 0.37 | 42 | 2.66 | 3.24 |
| 20 | 3.43 | 3.59 | 43 | 3.14 | 0.64 |
| 21 | 10.69 | 1.38 | 44 | 3.48 | 0.82 |
| 22 | 2.73 | 4.31 | 45 | 2.36 | 1.78 |
| 23 | 2.55 | 1.56 | 46 | 6.66 | 0.00 |

Literature Cited

1. Abdellah, F. G., and Levine, E., "Work Sampling Applied to the Study of Nursing Personnel," Nursing Research, Vol. 3, No. 1, June, 1954, pp. 11-16.
2. Abruzzi, A., "Determining Reliable Delay Allowances," Management Review, Vol. 37, No. 4, April, 1948, pp. 198-201.
3. Abruzzi, A., "Delay Allowances by Statistical Methods," Columbia Engineering Quarterly, Vol. 1, No. 1, May 1948, pp. 6-8, 23.
4. Abruzzi, A., "Applying the Control Chart Method to Worker Evaluation and Incentive Plans," Personnel, Vol. 25, No. 3, November, 1948, pp. 204-210.
5. Abruzzi, A., Work Measurement, Columbia University Press, New York, 1952, pp. 246-272.
6. Addoms, J., and Long, F. A., "Work Sampling in the Office" Office Management Series Report 140, American Management Association, 1955,
7. Agett, A. J., and Walker, W. D., "Corning Engineers Check on Selves," Plant Engineering, Vol. 10, No. 2, February, 1956, pp. 122-124.
8. Alenik, L., and Koepp, L., "Control Soaring Costs Through Work Sampling," Mill and Factory, Vol. 67, November, 1900, pp. 122-125.
9. Alford, L. P., and Bangs, J. R., Production Handbook, The Ronald Press Co., New York, 1944, pp. 514-519.
10. Allderige, J. M., "Statistical Sampling, a Sharp Tool for Industrial Engineers," Advanced Management, Vol. 16, No. 9, September, 1951, pp. 12-15.
11. Allderige, J. M., "Work Sampling Without Formulas," Factory Management and Maintenance, Vol. 112, No. 3, March 1954, pp. 136-138.
12. Allderige, John M., "Tackling Old Problems with New Tools from Operations Research" Advanced Management, Vol. 21, No. 9, September 1956, pp. 9-12.
13. Allderige, J. M., "Applying Mathematical Techniques to Industrial Engineering Problems," Proceedings Tenth Annual Industrial Engineering Institute, University of California, Los Angeles, California. February, 1958, pp. 1-8.

14. Allderige, J. M., "Comments on Work Sampling in an Engineering Design Department," The Journal of Industrial Engineering, Vol. XII, No. 3, May-June, 1961, pp. 179-181.
15. Allen, L. A., "Making the Foreman a Manager," Management Record, Vol. 16, No. 18, August, 1954, pp. 294-296.
16. Anderson, J. V., "Work Sampling Technique," American Society of Quality Control Transactions of the Mid-Atlantic Conference, 1956, pp. 53-61.
17. Barlament, R. L., "Photographic Work Sampling (Memo-Motion) New Cost Control Technique," Textile World, Vol. 108, September, 1958, pp. 52-53.
18. Barnes, R. M., "Work Sampling," Proceedings Twelfth Management Engineering Conference, S. A. M. - American Society of Mechanical Engineers, New York, April 1957.
19. Barnes, R. M., and Andrews, R. B., "Work Sampling," Proceedings Eighth Annual Industrial Engineering Institute, University of California, Los Angeles, Feb. 1956, pp. 9-10.
20. Barnes, R. M., "Measuring Work by Sampling Methods," Proceedings of Seventh Conference of American Institute of Industrial Engineers, Washington, D. C., May, 1956, pp. 6-1 to 6-21.
21. Barnes, R. M., Buffa, E. S., and Bary, J. A., "A Statistical Study of Performance Sampling in Work Measurement," A Booklet, University of California, Los Angeles, June, 1953 (out of print).
22. Barnes, R. M., Work Sampling, John Wiley and Sons, Inc., New York, 1956
23. Barnes, R. M., and Andrews, R. B., "Performance Sampling in Work Measurement," The Journal of Industrial Engineering, Vol. VI, No. 6, November - December, 1955, pp. 8-18.
24. Barnes, R. M., Buffa, E. S., Demangate, D. C., and Andrews, R. B., "Performance Sampling in Work Measurement," Proceedings Sixth Annual Industrial Engineering Institute, University of California, Los Angeles, February, 1954, pp. 85-89.
25. Barnes, R. M., Work Sampling, John Wiley and Sons, Inc. New York, New York, 1956.
26. Basu, S. K., and Dastidar, B. P. G., "Testing and Development of Activity Sampling in Indian Industry," Institution of Engineers (India) Journal, Vol. 38, No. 4, Part 2, December, 1957, pp. 353-367.
27. Bogenreef, C. A., "You Can Use Work Sampling this Way to Measure Indirect Work for Cost Control," Factory Management and Maintenance, Vol. 110, No. 12, December, 1952, pp. 113-114.

28. Breese, W. L., "Work Sampling an Aid to Cost Control," Journal of the Society of Automotive Engineers, Vol. 66, May, 1958, p. 107.
29. Brisley, C. L., "Work Sampling - A Cost Cutting Technique," Proceedings of the American Institute of Industrial Engineers, May 1955, pp. 168-179.
30. Brisley, C. L., "How You Can Put Work Sampling to Work," Factory Management and Maintenance, Vol. 110, No. 7, July 1952, pp. 84-89.
31. Brisley, C. L., "Ratio-Delay Developments," Proceedings Seventh Time Study and Methods Conference, S.A.M. - A.S.M.E., New York, 1951, pp. 162-192.
32. Brisley, C. L., "The How and Why of Ratio Delay Study," Proceedings Fifteenth Time and Motion Study and Management Clinic, Industrial Management Society, Chicago, Ill., November, 1951.
33. Brown, G. B., Smoothing, Forecasting, and Prediction of Discrete Time Series, Prentice Hall, 1963, p. 350.
34. Brownstein, H., "Make a Ratio-Delay Study to Control Labor Costs," Textile World, Vol. 107, October, 1957, pp. 102-103.
35. Bryant, B. M., "Work Sampling in Maintenance," Factory, Vol. 117, No. 12, December 1959, pp. 72-76.
36. Buckland, W. R., "A Review of the Literature of Systematic Sampling," Journal of the Royal Statistical Society, Series B, Vol. 13, 1951, pp. 208-215.
37. Chemical Week, "Keeping Tabs on Time: Ratio Delay," Vol. 76, March 5, 1955, pp. 46, 48, 52.
38. Cochran, W. G., "Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations," Annals of Mathematical Statistics, Vol. 17, 1946, pp. 164-177.
39. Cochran, W. G., Sampling Techniques, John Wiley and Sons, New York, 1953, pp. 160-187.
40. Cogan, A. J. and Stillian, G. N., "Ratio Delay Studies; A Quick New Way to Get Downtime Data," Factory Management and Maintenance, Vol. 110, No. 3, March, 1952, pp. 136-137.
41. Connor, R. J., "Hospital Work Sampling with Associated Measures of Production," The Journal of Industrial Engineering, Vol. XII, No. 2, March-April, 1961, pp. 105-108.

42. Conway, R. W., "Some Statistical Considerations in Work Sampling," The Journal of Industrial Engineering, Vol. VIII, No. 2, March-April, 1957, pp. 107-111.
43. Correll, D. S., and Barnes, R. M., "Industrial Application of the Ratio-Delay Method," Advanced Management in two parts, Vol. 15, No. 8, August, 1950, pp. 10-12, and Vol. 15, No. 9, September, 1950, pp. 15-18.
44. Cote, L. J., and Scott, B. J., "Comparison of All-Day Time Study with Work Sampling," Journal of Industrial Engineering, January-February, 1956, Vol. VII, No. 1, pp. 31-34.
45. Davidson, H. O., "Functions and Bases of Time Standards," American Institute of Industrial Engineers, Columbus, Ohio, 1952.
46. Davidson, H. O., "Activity Sampling and Analysis - Present State of Theory and Practice," A.S.M.E. Paper No. 53-F-24, 1953 (Paper not Published).
47. Davidson, H. O., Hines, W. W., and Newberry, T. L., "The Error of Estimate in Systematic Activity Sampling," The Journal of Industrial Engineering, Vol. XI, No. 4, July-August, 1960, pp. 290-292.
48. Davidson, H. O., "Work Sampling Eleven Fallacies," The Journal of Industrial Engineering, Vol. XI, No. 5, September-October, 1960, pp. 367-377.
49. Davis, H., "A Mathematical Evaluation of a Work Sampling Technique," Naval Research Logistics Quarterly, Vol. 2, No. 1, 2, March-June, 1955, pp. 111-117.
50. Deming, W. E., and Simmons, W. R., "On the Design of a Sample for Dealer Inventories," Journal of the American Statistical Association, Vol. 41, 1946, pp. 16-33.
51. Drui, A. B., "Industrial Research on the Effects of Pace on Work Sampling Results," The Journal of Industrial Engineering, Vol. XII, No. 1, January-February, 1961, pp. 28-31.
52. Dumene, R. L., "Work Sampling: A Statistical Control for Plan Maintenance and Construction," Middle Atlantic Conference Transactions of the American Society for Quality Control, Philadelphia, Pennsylvania, February 15-16, 1957, pp. 51-60.
53. Eastman, R. M., "Reducing Office Downtime with Ratio-Delay," Management Review, Vol. 42, September, 1953, pp. 524-525.
54. Factory Management and Maintenance, "New Look at Allowances," Vol. 113, August, 1955, pp. 112-113.

55. Finney, D. J., "Random and Systematic Sampling in Timber Surveys," Forestry, Vol. 22, 1948, pp. 1-36.
56. Finney, O. J., "An Example of Periodic Variation in Forest Sampling," Forestry, Vol. 23, 1950, pp. 96-111.
57. Fisher, R. A., and Mackenzie, W. A., "The Correlation of Weekly Rainfall," Quarterly Journal of the Royal Meteorological Society, Vol. 48, 1922, pp. 234-245.
58. Flowerdew, D. A. J., Malin, P. W., "Systematic Activity Sampling," Journal of Industrial Engineering, Vol. XIV, No. 41, July-Aug., 1963, p. 201.
59. French, C. E., "Fast and Simple Method for Checking Labor Performance," Food Engineering, Vol. 29, April, 1957, pp. 63-65.
60. Gautschi, W., "Some Remarks on Systematic Sampling," Annals of Mathematical Statistics, Vol. 28, 1957, pp. 285-295.
61. Gaylord, L. E., "Indirect Labor Standards Without Time Study," Proceedings of the Fourth Industrial Engineering Conference, (Technical Bulletin No. 37) West Virginia University, Morgantown, W. Virginia, 1953, pp. 73-79.
62. Gaylord, L. E. and Gillespie, C. R., "How to Establish Material Handling Incentive Standards without Time Studies," Flow, Vol. 9, No. 5, February, 1954, pp. 78-81.
63. Gustat, G. H., "Applications of Work Sampling Analysis," Proceedings of Tenth Time Study and Methods Conference, SAM-ASME, New York, April, 1955, pp. 130-143.
64. Haines, I. L., "Work Sampling by Fixed Interval Study," The Journal of Industrial Engineering, Vol. IX, No. 4, July-August, 1958, pp. 266-268.
65. Hajeck, J., "Optimum Strategy and Other Problems in Probability Sampling," Casopis Pest. Mat., Vol. 84, 1959, pp. 387-423.
66. Halley, E., "An Estimate of the Degrees of the Mortality of Mankind Drawn from Curious Tables of the Births and Funerals at the City of Breslau," Royal Society of London Philosophical Transactions, Abridged, 1693, p. 483-491.
67. Halsey, J. J., "A New Model for Work Sampling - The GREDS Theory," The Journal of Industrial Engineering, Vol. XI, No. 6, November-December, 1960, pp. 501-508.
68. Hansen, B. L., Work Sampling for Modern Management, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1960.

69. Hasel, A. A., "Sampling Error in Timber Surveys," Journal of Agricultural Research, Vol. 57, 1938, pp. 713-736.
70. Hatton, W. J., "Work Load by Sampling," Time and Motion Study, London, Vol. 5, No. 1, January, 1956.
71. Haynes, J. D., An Empirical Investigation of Sampling Methods for an Area, M. S. Thesis, University of North Carolina.
72. Heiland, R. E. and Richardson, W. J., "Work Sampling Works for a Small Office Staff," National Association of Cost Accountants Bulletin, Vol. 36, No. 9, May, 1955, pp. 1157-1167.
73. Heiland, R. E., and Richardson, W. J., Work Sampling, McGraw-Hill Book Company, Inc., New York, 1957.
74. Heller, R. P., "Work Sampling," Proceedings of the Ninth Annual Industrial Engineering Conference, West Virginia University Bulletin (TECHNICAL) No. 51, June 1958, pp. 74-77.
75. Homeyer, P. G., and Black, C. A., "Sampling Replicated Field Experiments on Oats for Yield Determinations," Soil Science Society Proceedings, Vol. 11, 1946, pp. 341-344.
76. Ireson, W. G., Grant, E. L. (ed.), Handbook of Industrial Engineering and Management, Prentice-Hall, Englewood Cliffs, N. J., 1955.
77. Isherwood, J. D., "Labor Cost Analysis by Work Sampling in the Small Business," The Journal of Industrial Engineering, Vol. XI, No. 5, Sept.-Oct., 1960, pp. 417-420.
78. Johnson, F. A., "A Statistical Study of Sampling Methods for Tree Nursery Inventories," Journal of Forestry, Vol. 41, 1943, pp. 674-689.
79. Jones, A. E., "Systematic Sampling of Continuous Parameter Populations," Biometrika, Vol. 35, 1948, pp. 283-290.
80. Jones, D., "Quick Way to Figure Machine Interference," Factory Management and Maintenance, Vol. 112, July, 1954, p. 132.
81. Jones, H. L., "Investigating the Properties of the Sample Mean by Employing Random Subsample Means," The Journal of the American Statistical Association, Vol. 51, 1956, pp. 54-83.
82. Jowett, G. H., "The Accuracy of Systematic Sampling from Conveyor Belts," Applied Statistics, Vol. 1, 1952, pp. 50-59.
83. Jaske, W. M., "Work Measurement by Work Sampling," Factory Management and Maintenance, Vol. 113, No. 1, January, 1955, pp. 102-106.

84. Kendall, M. G., "A Continuation of Dr. Jones' Paper," Biometrika, Vol. 35, 1948, pp. 291-296.
85. Kinniburgh, W., and McTaggart, G., "Snap Observation Helps Time Study in Building," The Manager, March, 1954, pp. 179-182.
86. Klein, M., "Double Sample Estimation in Work Measurement," The Journal of Industrial Engineering, Vol. X, No. 3, May-June, 1959, pp. 193-196.
87. Kruse, M. E., "Getting the Facts; Introduction to Work Sampling," Plant, Vol. 23, May, 1961, pp. 47-49.
88. Lifson, K. A., "Work Sampling; Uses and Abuses," Proceedings Eleventh National Conference of the American Institute of Industrial Engineers, May, 1960, pp. 79-80.
89. Leighty, W. R., "Work Sampling in Engineering," Proceedings Thirteenth Annual Industrial Engineering Institute, University of California, Los Angeles, February, 1961, pp. 42-48.
90. MacNiece, E. H., "Work Sampling: Newest Way to Check Maintenance Efficiency," Factory Management and Maintenance, Vol. 111, No. 7, July, 1953, pp. 110-112.
91. Macomber, Frank S., "Time for a Time-Utilization Survey?" Factory Management and Maintenance, Vol. 113, No. 5, May, 1955, pp. 88-91.
92. Madow, L. H., "Systematic Sampling and Its Relation to Other Sampling Designs," Journal of the American Statistical Association, Vol. 41, 1946, pp. 204-21.
93. Madow, W. G., and L. H., "On the Theory of Systematic Sampling, I," Annals of Mathematical Statistics, Vol. 15, 1944, pp. 1-24.
94. Madow, W. G., "On the Theory of Systematic Sampling, II," Annals of Mathematical Statistics, Vol. 20, 1949, pp. 333-354.
95. Madow, W. G., "On the Theory of Systematic Sampling, III," Annals of Mathematical Statistics, Vol. 24, 1953, pp. 101-106.
96. Maguire, J., "Work Sampling Sold Safety for Us," Factory Management and Maintenance, Vol. 113, February, 1955, pp. 92-93.
97. Mahaffey, D. E., "Work Sampling - To Set Standards on Indirect Labor," Factory Management and Maintenance, Vol. 112, No. 11, November, 1954, pp. 142-144.
98. Malcolm, D. G., and Sammet, "Work Sampling Applications," Journal of Industrial Engineering, Vol. V, No. 3, May-June, 1954, pp. 4-6, 23.

99. Management Review, "Studying a Job by Work Sampling," Vol. 18, No. 2, February, 1954, pp. 86-87.
100. Mandel, B. J., "A Short Cut to Fact Finding," Dun's Review and Modern Industry, Vol. 65, No. 2325, May, 1955, pp. 46, 56-58.
101. Mathis, L. H., "How to Use Time Sampling," The Office, Vol. 38, No. 5, November, 1953, pp. 89-90.
102. McAllister, G. E., Nelson, R., "Who Threw the Monkey-Wrench in Mr. Tippet's Theory," Industrial Logistics Research Project, Technical Report No. 5, University of California, Los Angeles, 1953.
103. McAllister, G. E., "Random Ratio Delay," Journal of Industrial Engineering, Vol. 4, No. 3, pp. 15-24, Aug. 1953.
104. McNaughton, J. B., "Work Sampling at Executive Level," Advanced Management, Vol. 21, October, 1956, pp. 12-13.
105. Millikin, J., "Random Sampling Boosts Labor Performance," American Machinist Metalworking Manufacturing, Vol. 104, November 28, 1960, pp. 148-150.
106. Mindlin, A., "The Application of Sample Survey Methodology to Work Sampling," The Journal of Industrial Engineering, Vol. X, No. 4, July-August, 1959, pp. 286-295.
107. Moder, J. J., Halladay, W. J., "Work Sampling Applied to Long Cycle Operations Performed by a Variable Labor Force," Journal of Industrial Engineering, Vol. VII, No. 4, July-August, 1956, pp. 164-167.
108. Morrow, R. L., "Ratio-Delay Study," Mechanical Engineering, Vol. 63, No. 4, April, 1941, pp. 302-303. (Abstract of Paper).
109. Morrow, R. L., Time Study and Motion Economy, Ronald Press. New York, 1946.
110. Morrow, R. L., Motion Economy and Work Measurement, Ronald Press Co., New York, 1957, pp. 273-319.
111. Moskowitz, A. D., "Two New Twists in Performance Rating Make Work Sampling ... Improve Productivity in Bench Assembly Jobs and...Reduce Manpower Requirements in Indirect Jobs," Factory, Vol. 120, No. 6, June, 1962, pp. 71-74.
112. Mueller, R. K., Effective Management through Probability Controls, Funk and Wagnalls Co., New York, New York, 1955, pp. 404-407.
113. Murray, M. A., "Standards for Clerical Work," Proceedings Twelfth Annual Industrial Engineering Institute, University of California, Los Angeles, February, 1960, pp. 49-53.

114. Nadler, G., and Denholm, D. H., "p-Charts Improve Methods for Determining Downtime," Iron Age, Vol. 170, No. 22, Nov. 1952, pp. 110-113.
115. Nadler, G., "What Do Your Foremen Do? - an Occurrence Study," Mill and Factory, Vol. 53, No. 4, October, 1953, pp. 88-92.
116. Nadler, G., Motion and Time Study, McGraw-Hill Book Co., New York, New York, 1955, pp. 476-484.
117. Niebel, B. W., "Work Sampling Gives Better Time Study Allowances," Tool Engineer, Vol. 35, No. 5, November, 1955, pp. 115-116.
118. Niebel, B. W., Motion and Time Study, Richard D. Irwin, Homewood, Illinois, 1955, pp. 334-342.
119. Nordsbog, A. W., and Crump, S. L., "Systematic and Random Sampling for Estimating Egg Production in Poultry," Biometrics, Vol. 4, 1948, pp. 223-233.
120. O'Neill, R. R., and Carrabino, J. D., "Work Sampling (Ratio-Delay) Statistical Techniques Applied to a Materials Handling System," Proceedings Sixth Annual Industrial Engineering Institute, University of California, Los Angeles, February, 1954, pp. 80-84.
121. Osborne, J. B., "Sampling Errors of Systematic and Random Surveys of Cover Type Areas," Journal of the American Statistical Association, Vol. 37, 1942, pp. 256-264.
122. Owens, B. W., "Sequential Work Sampling," Trend in Engineering at the University of Washington, Vol. VI, No. 1, January, 1954, pp. 23-27.
123. Page, R. N., "Now-Easy Way to Sample Work and Sure Way to Use Your Data," Factory, Vol. 116, September, 1958, pp. 168-180.
124. Paul, O. A., "Delay Ratio Factor Can be Applied in the Office," Advanced Management, Vol. 21, No. 11, November, 1956, pp. 25-26.
125. Petersen, D. N., "Labor Cost Control Through Performance Sampling," Proceedings Eleventh Annual Industrial Engineering Institute,
126. Petro, J. S., "Using Ratio Delay Studies to Set Allowances," Factory Management and Maintenance, Vol. 106, No. 10, October, 1948, pp. 92-94.
127. Proudfoot, M. J., "Sampling with Transverse Traverse Lines," Journal of the American Statistical Association, Vol. 37, 1942, pp. 265-270.

128. Provost, R. G., A Statistical Evaluation of Three Methods Used in Sampling Memomotion Film, (Unpublished Masters Thesis, Georgia Institute of Technology) Atlanta, Georgia, June, 1960.
129. Quenoville, M. H., "Problems in Plane Sampling," Annals of Mathematical Statistics, Vol. 20, 1949, pp. 355-375.
130. The Rand Corporation, A Million Random Digits with 100,000 Normal Deviates, Free Press, Glencoe, Illinois, 1955.
131. Reul, R. I., and Richardson, "Work Sampling; Key to Reducing the Cost of Materials Handling Activities," Modern Materials Handling, Vol. 14, April, 1959, pp. 88-92.
- 132.. Richardson, W. J., "Work Sampling Today," Factory, Vol. 117, No. 9, September, 1959, pp. 122-127.
133. Ritchey, J. A., "An Unusual Work Sampling Application," The Journal of Industrial Engineering, Vol. X, No. 6, November-December, 1959, pp. 450-455.
134. Rogers, N. K., A Study of Work-Time Distribution Characteristics and Their Relationship to Delay Time Distribution Characteristics for Several Operators During Similar Work Periods, M. S. Thesis, 1956, Georgia Institute of Technology.
135. Rosander, A. C., "Work Sampling Techniques," Industrial Quality Control, Vol. XVI, No. 12, June, 1960, pp. 11-18.
136. Rosander, A. C., Guterman, H. E., and McKeon, H. J., "Use of Work Sampling for Cost Analysis and Control," American Statistical Association Journal, Vol. 53, June, 1958, pp. 382-397.
137. Rowe, A. J., "The Work Sampling Technique," American Society of Mechanical Engineers Transactions, Vol. 76, 1954 (Paper No. 53-SA-55), pp. 331-335.
138. Rowe, A. J., "Relative versus Absolute Errors in Delay Measurement," British Management Review, Vol. 13, No. 1, London, January, 1955, pp. 60-64, (also Research Report 24, University of California, Los Angeles, 1953).
139. Sammet, L. L., "The Ratio-Delay Study," Proceedings of the Third Industrial Engineering Institute, University of California, 1951, pp. 54-58.
140. Sammet, L. L., and Hassler, J. B., "Use of the Ratio-Delay Method in Processing Plant Operations," Agricultural Economics Research, Vol. 3, No. 4, October, 1951, pp. 124-134.

141. Sammet, L. L., and Malcolm, D. G., "Work Sampling Studies: Guide to Analysis and Accurate Criteria," The Journal of Industrial Engineering, Vol. 5, No. 4, July, 1954, pp. 9-12.
142. Schaeffer, E. H., "Observation Ratios: A Short Cut to Time and Cost Analysis," Factory Management and Maintenance, Vol. 99, No. 7, July, 1941, pp. 58-59, 158.
143. Schmid, M. D., "Work Measurement Sampling Cracks Toughest Standard Data Nut," Mill and Factory, Vol. 67, July, 1960, pp. 100-101.
144. Smith, W. P., "Work Sampling, Fast Way to Get Facts on Handling," Factory Management and Maintenance, Vol. 111, No. 5, May, 1953, pp. 70-71.
145. Stephan, F. F., "History of the Uses of Modern Sampling Procedures," Journal of the American Statistical Association, Vol. 43, 1948, pp. 12-39.
146. Stewart, M., "The Work Sampling Method," The Office, October, 1954, p. 904.
147. Timmins, A. A., "Work Sampling," British Cast Iron Research Association, Journal, Vol. 9, No. 6, November, 1961, pp. 777-787.
148. Tippet, L. H. C., "Statistical Methods in Textile Research Uses of Binomial and Poisson Distribution, A Snap Reading Method of Making Time Studies of Machines and Operations in Factory Surveys," Journal of the Textile Institute Transactions, Vol. 26, Feb., 1935, pp. T13-T70.
149. Tippet, L. H. C., "The Study of Industrial Efficiency with Special Reference to the Cotton Industry," Journal of the Royal Statistical Society, Vol. 110, Part 2, 1947, pp. 108-122.
150. Tippet, L. H. C., "The Ratio Delay Technique," Time and Motion Study, Vol. 2, No. 5, London, May, 1953, pp. 10-19.
151. Underhill, J. S., "First-Line Supervisors - How Do They Spend Their Time?" Mechanical Engineering, Vol. 79, No. 4, pp. 353-355.
152. Van Eman, C. J., "Work Sampling Will Work in Your Plant," Hydrocarbon Process and Petroleum Refiner, Vol. 41, January, 1962, pp. 131-134.
153. Voris, J. W., "Random Observations: A Short Cut to Equipment Utilization," Systems and Procedures Quarterly, Vol. 7, No. 1, February, 1956, pp. 15-20.
154. Waddell, H. L., "Work Sampling - A New Tool to Help Cut Costs, Boosts Production Make Decisions," (editorial), Factory Management and Maintenance, Vol. 110, No. 7, July, 1952, p. 83.

155. Weir, J. T., "Work Sampling in a Nail Mill," Wire and Wire Products, Vol. 34, October, 1959, pp. 1336-1337.
156. Wheeler, D. L., "Try Time Lapse Photography as Work Sampling Technique," Plant, Vol. 23, December, 1961, pp. 19-21.
157. White, C. G., "Work Sampling is Procedural Trouble-Shooting," National Association of Cost Accountants Bulletin, Vol. 37, Section 1, No. 2, October, 1955, pp. 238-245.
158. Wilding-White, C. F. B., "Planning Work Sampling Studies for Efficient Sampling and Processing of Data," The Journal of Industrial Engineering, Vol. XII, No. 5, September-October, 1961, pp. 343-348.
159. Williams, J. A. C., "Work Sampling Techniques in Work Study," The Engineer, Vol. 203, No. 4744, London, February 15, 1957, pp. 245-249.
160. Williams, R. M., "The Variance of the Mean of Systematic Samples," Biometrika, Vol. 43, 1956, pp. 137-148.
161. Wolf, M., Riordan, D. P., "Work Sampling in an Engineering Design Department," The Journal of Industrial Engineering, Vol. XII, No. 3, May-June, 1961, pp. 175-179.
162. Wolff, H. S., "Modern Techniques for Time and Motion Study in Physiological Research," Ergonomics, August, 1959, pp. 354-362.
163. Wright, M. J., Improvement of Patient Care, G. P. Putnam's Sons, New York, 1954.
164. Yates, F., "Systematic Sampling," Philosophical Transactions of the Royal Society of London, Series A., Vol. 241, 1948, pp. 345-377.
165. Yates, F., Sampling Methods for Censuses and Surveys, 3rd Ed., Hafner Publishing Company, New York, 1960.

Other References

1. Allderige, J. M., "Designing the Measurement and Analysis of Work," American Society for Quality Control National Convention, Transactions, 1957, pp. 151-203.
2. Anderson, J. V., "Applications of the Work Sampling Technique," Proceedings First Conference on Administrative Applications, American Society for Quality Control, 1957.
3. Baley, G. V., and Hammersley, J. M., "The Effective Number of Independent Observations in an Autocorrelated Time Series," Journal of the Royal Statistical Society, Supplement, Vol. 8, 1946, pp. 184-197.
4. Barnes, R. M., Motion and Time Study, 4th Edition, John Wiley and Sons, Inc., New York, 1958.
5. Bartlett, M. S., "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series," Journal of the Royal Statistical Society, Supplement, Vol. 8, 1946, pp. 27-41.
6. Business Week, "Studying a Job - by Work Sampling," No. 1267, December 12, 1953, pp. 58-62.
7. Chantal, R., "Determination des Temps Improductifs Discontinues," Travail et Methodes, September 1954, pp. 21-27.
8. Connor, R. J., "A Work Sampling Study of Variations in Nursing Work Load," Hospitals May 1, 1961, pp. 40-41, 111.
9. Dalenius, T., Hajeck, J., and Zabrzycki, S., "On Plane Sampling and Related Geometrical Problems," Proceedings, Fourth Berkeley Symposium on Mathematical Statistics Probability, Vol. 1, University of California Press, 1961, pp. 125-150.
10. Dalenius, T., "Recent Advances in Sample Survey Theory and Methods," The Annals of Mathematical Statistics, Vol. 33, No. 2, June, 1962, p. 330.
11. Das, A. C., "Two Dimensional Systematic Sampling and the Associated Stratified and Random Sampling," Sankhya, Vol. 10, 1950, pp. 95-108.
12. Deming, W. G., Some Theory of Sampling, John Wiley and Sons Inc., New York, 1950.
13. Gomberg, W. A., "The Reliability of Time Study Standards," Management Review, February, 1948, pp. 83-84.

14. Gustat, G. H., "Incentives for Indirect Labor," Proceedings of the Fifth Industrial Engineering Institute, University of California, 1953, pp. 80-86.
15. Hansen, M. H., Hurwitz, W. N., and Madow, W. G., Sample Survey Methods and Theory, New York, John Wiley and Sons, Inc.
16. Heller, R. P., "Work Sampling," Proceedings of the Seventh Annual Industrial Engineering Conference, West Virginia University Bulletin (TECHNICAL) No. 43, June 1956, pp. 14-20.
17. Holmes, D. S., "Notes on Work Sampling Sample Size," The Journal of Industrial Engineering, Vol. IX, No. 4, July-August, 1958, pp. 242-243.
18. How to Study Nursing Activities in a Patient Unit, United States Government Printing Office, Washington, D. C., 1954.
19. Johnson, P. O., Rao, M. S., Modern Sampling Methods, University of Minnesota Press, Minneapolis, Minnesota, 1959.
20. Jones, H. L., "The Application of Sampling Procedures to Business Operations," Journal of the American Statistical Association, Vol. 50, 1955, pp. 763-774.
21. Karger, D. W., Bayha, F. H., Engineered Work Measurement, The Industrial Press, 93 Worth St., New York 13, New York, 1957.
22. Longdon, R. V., "Work Sampling Measures Job Efficiency," Electrical World, Vol. 151, May 11, 1959, p. 55.
23. MacNiece, E. H., "How Work Sampling Can Help You Rate Your Key People," Factory Management and Maintenance, Vol. 110, No. 10, October, 1952, pp. 98-100.
24. Mueller, R. K., Effective Management through Probability Controls, Funk and Wagnalls Co., New York, 1950, p. 3.
25. Ritchey, J. A., "Work Sampling for Measuring Clerical Work," The Office, Vol. 48, November, 1958, p. 96.
26. Solonim, M. J., Sampling in a Nutshell, Simon and Schuster, New York.
27. Torgersen, P. E., "An Example of Work Sampling in the Hospital," The Journal of Industrial Engineering, Vol. X, No. 3, May-June, 1959, pp. 197-200.
28. Whittle, P., "Tests of Fit in Time Series," Biometrika, Vol. 39, 1952, pp. 309-318.

29. Whittle, P., "The Analysis of Multiple Stationary Time Series," Journal of the Royal Statistical Society, Series B., Vol. 15, 1952, pp. 125-139.
30. Whittle, P., "On Stationary Processes in the Plane," Biometrika, Vol. 41, 1954, pp. 434-449.
31. Zubrycki, S., "Remarks on Random Stratified and Systematic Sampling in a Plane," Colloq. Math., Vol. 6, 1958, pp. 251-264.

VITA

William W. Hines was born in Tampa, Florida, on December 12, 1932, the son of Emmett and Willie Mae (nee Whaley) Hines. He attended public schools in Tampa, Florida, and Germantown, Tennessee, where he was graduated from high school in May, 1950.

In September of 1950, the author entered the University of Tennessee where he undertook a program of study in mechanical engineering. He was a member of Alpha Tau Omega social fraternity and recipient of the Lockett Engineering Award. In January of 1953, the author began residence and a program of study in mathematics at Memphis State University. He was a member of the Arnold Air Society, and he was graduated as a distinguished military graduate in March, 1954, with the degree Bachelor of Science (Mathematics).

On April 31, 1954, the author entered the United States Air Force as a second lieutenant. After spending several months in special training at the Wright-Patterson Air Force Base in Ohio, he was assigned to the Directorate of Maintenance Engineering at the Oklahoma City Air Material Area Headquarters. His duties were as Mathematical Officer. He was released from active duty on April 30, 1956.

Returning to civilian life, the author was employed by an engineering consulting firm in Memphis, Tennessee. In September, 1956, he enrolled in the Georgia Institute of Technology to pursue a program of study in industrial engineering. The author was employed by the

Department of Mathematics, where he taught on a part time basis. He received the degree Master of Science in Industrial Engineering in June, 1958. The topic of the master's thesis was: The Development and Application of Probability Distributions of Aircraft Engine Removals, and the thesis was written under the direction of Dr. Joseph J. Moder. The author was elected to the Society of Sigma Xi and Tau Beta Pi honorary societies.

In September, 1958, the author was enrolled in the Ph. D. program of the School of Industrial Engineering at the Georgia Institute of Technology, and in June, 1959, he began a full time teaching assignment as an instructor in the School of Industrial Engineering.

On August 22, 1959, the author was married to Barbara Gayle Fayssoux. A daughter, Jennifer Fayssoux, was born in November, 1960. In the spring of 1961, the author was promoted to the rank of Assistant Professor, and in July, 1961, he began a fifteen-month period during which he shared his working time between the Rich Electronic Computer Center and the School of Industrial Engineering.

In February of 1963, a son, William Whaley, was born. During the years 1962 through 1963, the author served as program chairman for the Atlanta Chapter of the American Institute of Industrial Engineers. He has been engaged as Assistant Editor and is now Technical Notes Editor of the Journal of Industrial Engineering.

The author was coauthor of the following publications during the period of study towards the doctor of philosophy degree:

"Procedures for Forecasting Aircraft Engine Maintenance Removal Quantities," Journal of Industrial Engineering, November-December 1960, pp. 471-477. (with Joseph J. Moder)

"The Error of Estimate in Systematic Activity Sampling," Journal of Industrial Engineering, July-August 1960, pp. 290-292. (with H. O. Davidson and T. L. Newberry)